KOMPARASI METODE DATA MINING *K-NEAREST NEIGHBOR* DENGAN *NAÏVE BAYES* UNTUK KLASIFIKASI KUALITAS AIR BERSIH
(Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:
Maulana Aditya Rahman
NIM: 145150201111029

PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2018

PENGESAHAN

KOMPARASI METODE DATA MINING *K-NEAREST NEIGHBOR* DENGAN *NAÏVE BAYES* UNTUK KLASIFIKASI KUALITAS AIR BERSIH
(Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh :
Maulana Aditya Rahman
NIM: 145150201111029

Skripsi ini telah diuji dan dinyatakan lulus pada
24 Juli 2018

Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I	Dosen Pembimbing II
Nurul Hidayat, S.Pd, M.Sc	Dr.EngAhmad Afif S, S.Si, M.Kom
NIP. 196804302002121001	NIK. 2012018206231001

Mengetahui
Ketua Jurusan Teknik Informatika

Tri Astoto Kurniawan, S.T., M.T., Ph.D
NIP. 197105182003121001

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 01 Juli 2018

Maulana Aditya Rahman
NIM 145150201111029

KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia yang telah diberikan-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "**KOMPARASI METODE DATA MINING *K-NEAREST NEIGHBO*R DENGAN *NAÏVE BAYES* UNTUK KLASIFIKASI KUALITAS AIR BERSIH (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)**" dapat diselesaikan tepat waktu. Tidak pernah lupa sholawat serta salam juga ditujukan kepada Rasulullah, Nabi Muhammad SAW dan para sahabat.

Skripsi ini memiliki tujuan sebagai tugas akhir penulis selama mengikuti masa perkuliahan dan juga sebagai salah satu syarat untuk mendapat gelar Sarjana Komputer dari Fakultas Ilmu Komputer, Universitas Brawijaya, Malang.

Penyelesain skripsi ini tentunya tak pernah terlepas dari bimbingan dan bantuan yang melibatkan banyak pihak. Terlepas dari hal tersebut penulis ingin mengungkapkan rasa terima kasih yang sebesar-besarnya kepada :

1. Bapak Nurul Hidayat, S.Pd, M.Sc selaku dosen pembimbing pertama yang telah meluangkan waktunya untuk membimbing dan membantu penulis dalam penyusunan dan pengerjaan skripsi.
2. Bapak Dr.EngAhmad Afif Supianto, S.Si, M.Kom selaku dosen pembimbing kedua yang telah meluangkan waktunya untuk membimbing dan membantu penulis dalam penyusunan dan pengerjaan skripsi.
3. Bapak Aries Yuswantono selaku Direktur Utama PDAM Tirta Kencana Kabupaten Jombang yang telah memberi izin kepada penulis untuk melakukan penelitian di tempat tersebut.
4. Bapak Wayan Firdaus Mahmudy, S.Si, M.T, Ph.D selaku Dekan di Fakultas Ilmu Komputer.
5. Bapak Tri Astoto Kurniawan, S.T, M.T, Ph.D selaku Ketua Jurusan Teknik Informatika di Fakultas Ilmu Komputer.
6. Bapak Agus Wahyu Widodo, S.T, M.Cs selaku ketua Program Studi Teknik Informatika di Fakultas Ilmu Komputer.
7. Bapak dan Ibu Dosen Fakultas Ilmu Komputer yang selama masa perkuliahan bersedia memberikan ilmunya.
8. Seluruh *civitas* kemahasiswaan dan akademik terutama Mas Hermawan Dwi Putra yang telah banyak memotivasi dan memberi banyak pengalaman selama menjadi mahasiswa.
9. Keluarga penulis terutama kedua orang tua, Bapak Siswoto, S.H dan Ibu Betty Yuspitasari, S.Sos, M.Si dan saudari kandung Shinta Devy Permatasari yang selalu memberi doa, motivasi dan nasehat serta dukungan berupa moral dan materil kepada penulis.
10. Sephia Aldiariza Hernandha, selaku kekasih penulis yang selalu mendoakan, memotivasi dan memndukung penulis untuk menyelesaikan skripsi ini.

11. Adhy, Kamal, Bambang, Sendi, Josua, Iqbal, dan Danang, selaku sahabat dan keluarga kedua penulis yang selalu ada ketika penulis mengalami kesusahan dan menyerah dalam pengerjaan skripsi.
12. Teman-teman BIOS Exalt, Aziz, Silvi, Aya, Natasya, Ody, Aghni, Winny, Andhika dan yang lainnya yang telah memberi pengalaman yang sangat berharga ketika penulis berorganisasi dalam satu periode.
13. Eka, Nisa, Ninda, Sarah, Isradi, Elha dan teman-teman kelas C yang tidak bisa disebutkan penulis satu persatu yang telah membantu penulis beradaptasi dengan lingkungan kuliah pada saat masa awal perkuliahan.
14. Rhiezky, Ivan Prim, Juan, Farhan, Arum, Dea, Tipang dan teman-teman lainnya dari keluarga futsal dan sepak bola Fakultas Ilmu Komputer.
15. Elsa, Frisa, dan Shabrina selaku teman seperjuangan penulis dari masa SMA hingga berkuliah di Malang yang membantu penulis beradaptasi dengan lingkungan baru di Kota Malang.
16. Keluarga besar Ikatan Mahasiswa Jombang Universitas Brawijaya yang telah menerima dan menyambut penulis ketika pertama kali datang sebagai mahasiswa.
17. Semua pihak yang membantu penulis dalam penyelesaian skipsi yang tidak bisa disebutkan satu persatu

Penulis menyadari bahwa penyusunan skripsi ini jauh dari kata sempurna yang tak lepas dari kekurangan dan kesalahan, sehingga saran dan kritik yang membangun sangat diharapkan oleh penulis. Akhir kata, semoga skripsi ini bermanfaat bagi semua pihak yang menggunakannya. Terima Kasih.

Malang, 01 Juli 2018

Penulis
Adityarc9@gmail.com

ABSTRAK

Air adalah merupakan senyawa kimia yang sangat dibutuhkan bagi kelangsungan hidup makhluk hidup yang ada di bumi. Wilayah terluas di planet bumi merupakan air yang menutupi hampir 71% wilayah yang ada di bumi. Air juga merupakan zat yang sangat penting yang ada di bumi yang sangat dibutuhkan oleh semua makhluk hidup mulai dari tumbuhan, hewan dan manusia. Tumbuhan memerlukan air sebagai salah satu senyawa dalam proses fotosintesis. Hewan membutuhkan air untuk proses pencernaan makanan dan sebagai tempat tinggal. Air juga dibutuhkan manusia untuk keperluan sehari-hari seperti memasak, mencuci, mandi dan lain-lainnya. Air juga rentan terkontaminasi oleh bakteri-bakteri dan zat mineral yang berbahaya bagi tubuh manusia. Hal tersebut bisa terjadi dikarenakan tercemarnya sumber air atau tercemarnya lingkungan di sekitar sumber air. Dibutuhkan pengawasan dan pengolahan lingkungan sekitar sumber air sehingga dapat menghasilkan kualitas air yang bersih sesuai dengan standar kualitas air bersih dan memenuhi standar air yang layak dikonsumsi oleh manusia. Untuk menentukan klasifikasi kualitas air bersih terdapat banyak metode yang dapat digunakan. Untuk memilih metode klasifikasi yang paling cocok, dapat dilakukan komparasi antara beberapa metode. Penelitian ini melakukan komparasi antara metode *K-Nearest Neighbor* dan *Naïve Bayes*. Berdasarkan dari beberapa penelitian sebelumnya, metode *K-Nearest Neighbor* dan *Naïve Bayes* merupakan metode yang cukup baik dan menghasilkan tingkat akurasi yang cukup tinggi. Berdasarkan hasil pengujian, diperoleh rata-rata nilai akurasi metode *K-Nearest Neighbor* sebesar 82.42% dan rata-rata nilai akurasi metode *Naïve Bayes* sebesar 70.32%. Dapat disimpulkan bahwa metode yang paling baik untuk klasifikasi kualitas air bersih adalah metode *K-Nearest Neighbor*.

Kata kunci: kualitas air bersih, data mining, klasifikasi, komparasi metode, *k-nearest neighbor*, *naïve bayes*

ABSTRACT

Water is a chemical compound that is needed for the survival of living things on earth. The widest area on planet earth is water that covers almost 71% of the region on earth. Water is also a very important substance on earth that is needed by all living things from plants, animals and humans. Plants require water as one of the compounds in the process of photosynthesis. Animals need water for food digestion and as a place to live. Water is also needed humans for everyday purposes such as cooking, washing, bathing and others. Water is also vulnerable to contamination by bacteria and mineral substances that are harmful to the human body. This can happen due to pollution of water sources or contaminated environment around the water source. It takes the supervision and processing of the environment around the water source so as to produce clean water quality in accordance with the standard of clean water quality and meet the standard of water that is suitable for human consumption. To determine the classification of clean water quality there are many methods that can be used. To choose the best classification method, it can be comparated between several methods. This study comparing the K-Nearest Neighbor and Naïve Bayes methods. Based on several studies, the K-Nearest Neighbor and Naïve Bayes methods are quite good and yield a high degree of accuracy. Based on the test result, the average accuracy value of K-Nearest Neighbor method is 82.42% and the average accuracy of Naïve Bayes method is 70.32%. It can be concluded that the best method for water quality classification is K-Nearest Neighbor method.

Keywords: water quality, data mining, classification, comparison method, k-nearest neighbor, naïve bayes.

DAFTAR ISI

PENGESAHAN ii
PERNYATAAN ORISINALITAS iii
KATA PENGANTAR iv
ABSTRAK vi
Abstract vii
DAFTAR ISI viii
DAFTAR TABEL xi
DAFTAR GAMBAR xii
DAFTAR LAMPIRAN xiii
BAB 1 PENDAHULUAN 1
1.1 Latar belakang 1
1.2 Rumusan masalah 2
1.3 Tujuan 3
1.4 Manfaat 3
1.5 Batasan masalah 3
1.6 Sistematika pembahasan 3
BAB 2 LANDASAN KEPUSTAKAAN 5
2.1 Kajian Pustaka 5
2.2 Air 5
2.2.1 Pengertian Air 5
2.2.2 Pengertian Air Bersih 6
2.2.3 Pemanfaatan Air 6
2.3 Klasifikasi 7
2.4 Algoritme *Naïve Bayes* 7
2.5 Algoritme *K-Nearest Neighbor* 8
2.6 Pengujian Akurasi 8
BAB 3 METODOLOGI 9
3.1 Studi Pustaka 9
3.2 Pengumpulan Data 10
3.3 Implementasi 10

3.4 Lokasi Penelitian 10
3.5 Jadwal Kegiatan 10
BAB 4 Perancangan dan Implementasi sistem 11
4.1 Perancangan Sistem 11
4.1.1 Klasifikasi Menggunakan Metode *K-Nearest Neighbor* 12
4.1.2 Klasifikasi Menggunakan Metode *Naïve Bayes* 13
4.1.3 Proses Perhitungan Semua Data 14
4.2 Perancangan Pengujian 15
4.2.1 Pengujian Berdasarkan Nilai *K* pada Metode *K-Nearest Neighbor* 15
4.2.2 Pengujian Berdasarkan Rasio Perbandingan atau *Percentage Split* 15
4.2.3 Pengujian Berdasarkan Jumlah Data *Training* 16
4.3 Perancangan Algoritme 16
4.3.1 Perhitungan dengan *K-Nearest Neighbor* 18
4.3.2 Perhitungan dengan *Naïve Bayes* 26
4.4 Perancangan Antarmuka 28
4.5 Implementasi Sistem 29
4.5.1 Spesifikasi Perangkat Lunak 29
4.5.2 Spesifikasi Perangkat Keras 29
4.6 Implementasi Algoritme 30
4.7 Implementasi Tampilan Antarmuka 37
BAB 5 Pengujian dan Analisis 38
5.1 Hasil Pengujian 38
5.1.1 Hasil Pengujian Berdasarkan Nilai Atribut *K* pada Metode *K-Nearest Neighbor* 38
5.1.2 Hasil Pengujian Berdasarkan Rasio Perbandingan atau *Percentage Split* 40
5.1.3 Hasil Pengujian Berdasarkan Jumlah Data *Training* 41
5.2 Analisis Hasil 43
5.2.1 Analisis Hasil Pengujian Berdasarkan Nilai Atribut *K* pada Metode *K-Nearest Neighbor* 43

5.2.2 Analisis Hasil Pengujian Berdasarkan Rasio Perbandingan atau *Percentage Split* 43
5.2.3 Analisis Hasil Pengujian Berdasarkan Jumlah Data *Training* 45
BAB 6 Penutup 46
6.1 Kesimpulan..... 46
6.2 Saran 47
DAFTAR PUSTAKA..... 48
LAMPIRAN 49

DAFTAR TABEL

Tabel 3.1 Tabel Jadwal Kegiatan 10
Tabel 4.1 Contoh Tabel Pengujian Berdasarkan Nilai Atribut K..... 15
Tabel 4.2 Contoh Tabel Pengujian Berdasarkan Rasio Perbandingan 15
Tabel 4.3 Contoh Tabel Pengujian Berdasarkan Data Training..... 16
Tabel 4.4 Contoh Kasus Data Training Perhitungan Manual 16
Tabel 4.5 Contoh Kasus Dara Testing Perhitungan Manual..... 18
Tabel 4.6 Hasil Perhitungan Jarak Euclidean..... 18
Tabel 4.7 Hasil pengurutan jarak Euclidean..... 22
Tabel 4.8 Hasil keputusan berdasarkan nilai K 26
Tabel 4.9 Perhitungan Proabilitas Prior 26
Tabel 4.10 Perhitungan Probabilitas Likelihood Tidak Sesuai 26
Tabel 4.11 Perhitungan Probabilitas Likelihood Sesuai 27
Tabel 4.12 Hasil Perhitungan Probabilitas Posterior 28
Tabel 4.13 Spesifikasi Perangkat Lunak atau Software..... 29
Tabel 4.14 Spesifikasi Peangkat Keras atau Hardware 30
Tabel 4.15 Kode Program Implementasi Algoritme..... 30
Tabel 5.1 Tabel Hasil Pengujian Berdasarkan Nilai Atribut K pada Metode K-Nearest Neighbor 38
Tabel 5.2 Tabel Hasil Pengujian Berdasarkan Rasio Perbandingan 40
Tabel 5.3 Hasil Pengujian Berdasarkan Variasi Jumlah Data Training 42

DAFTAR GAMBAR

Gambar 3.1 Diagram metode tahapan penelitian 9
Gambar 4.1 Diagram Perancangan Klasifikasi Kualitas Air Bersih 11
Gambar 4.2 Diagram alir klasifikasi K-Nearest Neighbor..... 12
Gambar 4.3 Diagram alir klasifikasi Naïve Bayes 13
Gambar 4.4 Proses Perhitungan Semua Data..... 14
Gambar 4.5 Perancangan Tampilan Antarmuka Hasil 29
Gambar 4.6 Implementasi Tampilan Sistem..... 37
Gambar 5.1 Grafik Hasil Pengujian Berdasarkan Nilai Atribut K pada Metode K-Nearest Neighbor..... 43
Gambar 5.2 Grafik Hasil Pengujian Berdasarkan Rasio Perbandingan 44
Gambar 5.3 Grafik Hasil Pengujian Berdasarkan Variasi Jumlah Data Training ... 45

DAFTAR LAMPIRAN

Dataset Kualifikasi Air Bersih 49

BAB 1 PENDAHULUAN

1.1 Latar belakang

Air adalah merupakan senyawa kimia yang sangat dibutuhkan bagi kelangsungan hidup makhluk hidup yang ada di bumi. Wilayah terluas di planet bumi merupakan air yang menutupi hampir 71% wilayah yang ada di bumi. Di bumi ketersediaan volume air sebesar 1,4 triliun kilometer kubik. Terdapat banyak sumber air yang ada di bumi seperti air laut, air sungai, air permukaan, air atmosfir, air rawa dan air tanah.

Air juga merupakan zat yang sangat penting yang ada di bumi yang sangat dibutuhkan oleh semua makhluk hidup mulai dari tumbuhan, hewan dan manusia. Tumbuhan memerlukan air sebagai salah satu senyawa dalam proses fotosintesis. Hewan membutuhkan air untuk proses pencernaan makanan dan sebagai tempat tinggal. Air juga dibutuhkan manusia untuk keperluan sehari-hari seperti memasak, mencuci, mandi dan lain-lainnya. Oleh karena itu air sering disebut sebagai sumber kehidupan yang mana disitu ada air maka disitu pula terdapat kehidupan.

Air merupakan senyawa kompleks karena mengandung zat-zat dan mineral-mineral di dalamnya. Namun tidak semua zat dan mineral yang terkandung di air dapat dicerna dan diterima dengan baik oleh tubuh manusia. Air juga rentan terkontaminasi oleh bakteri-bakteri dan zat mineral yang berbahaya bagi tubuh manusia. Hal tersebut bisa terjadi dikarenakan tercemarnya sumber air atau tercemarnya lingkungan di sekitar sumber air. Begitu pentingnya peranan air dalam kehidupan sehingga dapat dinyatakan bahwa kualitas air dapat digunakan sebagai indikator tingkat kesehatan manusia(Situmorang, 2017). Dibutuhkan pengawasan dan pengolahan lingkungan sekitar sumber air sehingga dapat menghasilkan kualitas air yang bersih sesuai dengan standar kualitas air bersih dan memenuhi standar air yang layak dikonsumsi oleh manusia. Oleh karena itu perlu adanya upaya untuk menjaga kualitas dengan melakukan pemantauan dan pengukuran kualitas air.

Untuk mengetahui bahwa air tersebut memeliki kualitas yang sesuai syarat kesehatan dapat diketahui dari zat-zat atau mineral yang terkandung didalamnya. Namun dalam penentuan kualitas air masih menggunakan perhitungan manual seperti *Water Quality Index (WQI)* dan *STORET.* Pada metode *STORET* masih menggunakan metode penghitungan secara manual dengan menghitung satu-persatu data parameter. Kelemahan dari metode ini membutuhkan waktu yang cukup lama yaitu 1 sampai 30 hari sesuai dengan parameter apa yang diukur dan diteliti dan biaya yang digunakan cukup mahal. Sehingga, dalam mengatasi permasalahan klasifikasi terhadap kualitas air, peneliti mengusulkan penggunaan metode klasifikasi data dan dapat memberikan solusi dalam membantu proses penetuan terhadap klasifikasi kualitas air yang lebih efektif dan efisien.

Untuk menentukan klasifikasi kualitas air bersih terdapat banyak metode yang dapat digunakan seperti *K-Nearest Neighbor*, *Naïve Bayes*, *K-Means*,

Decision Tree dan lain sebagainya. Namun untuk memilih metode yang paling cocok, dapat dilakukan komparasi antara beberapa metode. Dalam penelitian sebelumnya oleh Rifwan Hamidi dengan menggunakan metode *Learning Vector Quantization* untuk klasifikasi kualitas air sungai menggunakan parameter sejumlah 7 masukan dengan menghasilkan akurasi sebesar 81.13%. Penelitian lain yang dilakukan oleh Mila Listiana pada tahun 2015 dalam kasus identifikasi tumbuh kembang anak balita dengan perbandingan algoritme *Decision Tree(C4.5)* dan *Naïve Bayes*, diperoleh hasil pengujian rata-rata nilai akurasi metode *Naïve Bayes* sebesar 96,89% dan algoritme *Decision Tree(C4.5)* sebesar 89,78%.

Penelitian lain yang dilakukan oleh Khafiizh Hastuti pada tahun 2012 pada prediksi data mahasiswa nonaktif dengan melakukan perbandingan metode klasifikasi *Logistic Regression, Decision Tree, Neural Network* dan *Naïve Bayes*. Penelitian tersebut menghasilkan tingkat akurasi metode *Logistic Regression* sebesar 81,64%, metode *Decision Tree* sebesar 95,29%, metode *Neural Network* sebesar 94,56% dan metode *Naïve Bayes* sebesar 93,47%. Selain itu terdapat penelitian lain yang dilakukan oleh Santoso pada tahun 2016 dengan melakukan komparasi metode *K-Nearest Neighbor* dan *Learning Vector Quantization (LVQ)* dengan studi kasus peramalan klasifikasi tingkat kemiskinan menghasilkan akurasi metode *K-Nearest Neighbor* sebesar 93.52% dan *Learning Vector Quantization (LVQ)* sebesar 75.93%.

Pada penelitian sebelumnya terdapat perbedaan tingkat akurasi beberapa metode untuk klasifikasi. Berdasarkan dari beberapa penelitian sebelumnya, metode *K-Nearest Neighbor* dan *Naïve Bayes* merupakan metode yang memiliki akurasi cukup tinggi. Untuk itu penelitian ini, peneliti akan melakukan komparasi antara metode *K-Nearest Neighbor* dan *metode Naïve Bayes* untuk mengetahui metode mana yang lebih baik dalam membantu melakukan klasifikasi terhadap kualitas air bersih. Berdasarkan uraian latar belakang, maka judul yang diambil dalam skripsi ini adalah "Komparasi Metode Data Mining *K-Nearest Neighbor* dan *Naïve Bayes* untuk Klasifikasi Kualitas Air Bersih".

1.2 Rumusan masalah

Berdasarkan latar belakang yang telah dijelaskan, dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana mengimplementasikan komparasi metode *K-Nearest Neighbor* dengan *Naïve Bayes* untuk klasifikasi kualitas air bersih ?
2. Berapa hasil tingkat akurasi sistem untuk klasifikasi kualitas air bersih menggunakan metode *K-Nearest Neighbor* dengan *Naïve Bayes* ?
3. Bagaimana pengaruh nilai *K*, rasio perbandingan data *training* dan data *testing*, dan variasi jumlah data *training* terhadap akurasi metode *K-Nearest Neighbor* dan *Naïve Bayes*?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah dirumuskan, dapat dirumuskan pula tujuan dari penelitian adalah sebagai berikut :

1. Mengimplementasikan komparasi metode *K-Nearest Neighbor* dengan *Naïve Bayes* klasifikasi kualitas air bersih.
2. Menguji tingkat akurasi sistem untuk klasifikasi kualitas air bersih menggunakan metode *K-Nearest Neighbor* dengan *Naïve Bayes.*
3. Mengetahui pengaruh nilai *K*, rasio perbandingan data *training* dan data *testing* dan variasi jumlah data *training* terhadap akurasi metode *K-Nearest Neighbor* dan *Naïve Bayes.*

1.4 Manfaat

Manfaat yang bisa diharapkan dari penelitian ini adalah membantu pihak PDAM Tirta Kencana untuk melakukan klasifikasi kualitas air dengan efektif agar tidak memakan waktu yang cukup lama dan biaya yang tinggi. Manfaat lain adalah untuk mengetahui metode mana yang cocok antara *K-Nearest Neighbor* dan *Naive Bayes* untuk klasifikasi kualitas air

1.5 Batasan masalah

Untuk memfokuskan penelitian yang akan dilakukan, maka permasalahan ini dibatasi oleh hal-hal sebagai berikut :

a. Data yang digunakan adalah data yang diperoleh dari PDAM Tirta Kencana Kabupaten Jombang dalam jangka waktu tahun 2016 hingga akhir tahun 2017
b. Data yang digunakan sejumlah 167 data
c. Parameter yang digunakan adalah bakteri *Coliform*, bakteri *Escherichia coli*, Mangan, TDS (*Total Dissolve Solid*) dan Khlorida.

1.6 Sistematika pembahasan

Sistematika penulisan penelitian pada penelitian skripsi ini akan diuraikan secara singkat sebagai berikut:

BAB 1 Pendahuluan

Berisi tentang latar belakang yang mendasari penelitian, rumusan masalah yang akan dibahas, tujuan penelitain, manfaat yang akan didapat setelah melakukan penelitian, batasan-batasan masalah dan sistematika pembahasan mengenai komparasi metode data mining *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi kualitas air bersih.

BAB 2 Landasan Kepustakaan

Berisi pembahasan tentang referensi serta literatur mengenai dasar teori tentang kualitas air bersih dan metode *K-Nearest Neighbor* dan *Naïve Bayes* yang digunakan untuk membangun sistem klasifikasi kualitas air bersih.

BAB 3 Metodologi

Berisi tentang langkah-langkah dari tahap awal penelitian hingga akhir penelitian dan perancangan algoritme yang akan digunakan dalam penelitian.

BAB 4 Perancangan dan Implementasi

Berisi tentang tahapan perancangan antarmuka, perancangan pengujian dan perancangan algoritme yang dilakukan dalam penelitian komparasi metode data mining *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi kualitas air bersih. Serta perhitungan manual metode *K-Nearest Neighbor* dan metode *Naïve Bayes* dan berisi tentang implementasi metode *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi kualitas air bersih berupa *source code* dan implementasi antarmuka sesuai dengan perancangan yang telah dibuat sebelumnya.

BAB 6 Pengujian dan Analisis

Berisi pembahasan tentang pengujian yang telah dilakukan pada sistem klasifikasi kualitas air bersih menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes* dan hasil analisis dari sistem tersebut.

BAB 7 Penutup

Pada bab ini berisi tentang kesimpulan dari penelitian yang telah dilakukan dan saran dari hasil yang telah dicapai.

BAB 2 LANDASAN KEPUSTAKAAN

2.1 Kajian Pustaka

Penelitian tentang perbandingan metode sudah sering dilakukan. Kajian terhadap penelitian tersebut sangat beragam sesuai dengan permasalahan yang diamati. Kajian pustaka pada penelitian ini adalah membandingkan penelitian ini dengan penelitian sebelumnya yang berjudul "Perbandingan Algoeitme *Decision Tree (C4.5)* dan *Naïve Bayes* pada Data Mining untuk Identifikasi Tumbuh Kembang Anak Balita" yang dilakukan oleh Mila Listiana pada tahun 2015. Penelitian tersebut menghasilkan tingkat akurasi algortime *Decision Tree* (C4.5) sebesar 89,78% dan metode *Naïve Bayes* sebesar 96,89%. (Listiana, 2015). Selain itu terdapat penelitian oleh Rifwan Hamidi dkk dengan judul "Implementasi *Learning Vector Quantization(LVQ)* untuk Klasifikasi Kualitas Air Sungai" menghasilkan akurasi sebesar 81.13% dengan menggunakan 7 parameter masukan.(Hamidi,2017).

Penelitian yang dilakukan oleh Santoso pada tahun 2016 yang berjudul "Perbandingan Metode K-Nearest Neighbor dan Learning Vector Quantization (LVQ) untuk Permasalahan Klasifikasi Tingkat Kemiskinan". Hasil pengujian menunjukkan bahwa rata-rata akurasi metode *K-Nearest Neighbor* adalah 93.52%, sedangkan metode *Learning Vector Quantization (LVQ)* sebesar 75,93%.(Santoso,2016).

Penelitian lain pada tahun 2012 dengan judul "Analisis Komparasi Algoritme Klasifikasi *Data Mining* untuk Prediksi Mahasiswa Non Aktif" oleh Khafiizh Hastuti menggunakan banyak metode dengan hasil akurasi yang beragam. Metode tersebut antara lain metode *Logistic Regression* dengan akurasi sebesar 81,64%, metode *Decision Tree* sebesar 95,29%, metode *Neural Network* sebesar 94,56% dan metode *Naïve Bayes* sebesar 93,47%.(Hastuti,2012)

Meskipun penelitian tentang komparasi atau perbandingan metode sudah sering dilakukan, namun penelitian tersebut masih layak dilakukan. Karena masih banyak metode yang perlu dibandingkan untuk mengetahui metode mana yang paling akurat oleh sebab itu penulis mengambil topik komparasi metode menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*.

2.2 Air

2.2.1 Pengertian Air

Air merupakan suatu zat cair yang terdiri dari hidrogen dan oksigen dengan rumus kimia H_2O yang tidak memiliki rasa, bau dan warna. Air memiliki sifat yang hampir dapat digunakan untuk keperluan apa saja. Hal tersebut membuat air tidak dapat dipisahkan dari semua bentuk kehidupan sampai saat ini selain energi dari matahari. Air adalah semua air yang terdapat pada diatas maupun di bawah permukaan tanah. Air dalam pengertian ini termasuk air permukaan, air tanah, air hujan dan air laut yang dimanfaatkan di darat.

Air dapat berupa air tawar dan air asin (air laut) yang merupakan bagian terbesar di bumi ini. Di dalam lingkungan alam proses, perubahan wujud,

gerakan aliran air (di permukaaan tanah, di dalam tanah, dan di udara) dan jenis air mengikuti suatu siklus keseimbangan dan dikenal dengan istilah siklus hidrologi (Robert J. Kodoatie, 2010).

2.2.2 Pengertian Air Bersih

Berdasarkan Peraturan Pemerintah Republik Indonesia Nomor 82 tahun 2011 tentang Pengelolaan Kualitas dan Pengendalian Pencemaran Air, air bersih adalah air yang dipergunakan untuk keperluan sehari-hari dan kualitasnya memenuhi persyaratan kesehatan air bersih sesuai dengan peraturan perundang-undangan yang berlaku dan dapat diminum apabila dimasak. Air bersih memiliki persyaratan antara lain syarat fisik, syarat kimiawi, dan syarat mikrobbiologi. Berdasarkan Peraturan Menteri Kesehatan Nomor 492 tahun 2010 tentang persyaratan kualitas air, batas syarat air dikatakan bersih dapat dilihat pada Tabel 2.1.

Tabel 2.1 Batas Syarat Air Bersih

No	Parameter	Kadar Maksimum	Satuan
1	Total Bakteri *Coliform*	< 0	Jumlah per 100 ml
2	*Escherichia Coli*	< 0	Jumlah per 100 ml
3	Mangan	< 0,4	Mg/Liter
4	*Total Disolved Solids* (TDS)	< 500	Mg/Liter
5	Khlorida	< 250	Mg/Liter

Syarat Fisik :

a. Tampilan jernih dan tidak berkeruh
b. Tidak berwarna apapun
c. Tidak berasa apapun
d. Tiadk berbau apapun
e. Suhu antara 10-25 derajat *Celcius*
f. Tidak menghasilkan endapan

Syarat Kimiawi :

g. Tidak mengandung bahan kimiawi yang beracun
h. Tidak mengandung bahan kimiawi yang berlebihan
i. pH air antara 6,5-9,2

Syarat Mikrobiologi :

a. Tidak mengandung kuman-kuman penyakit dan bakteri berbahaya seperti bakteri *Coliform* dan *Eschericia Coli*.

2.2.3 Pemanfaatan Air

Air sangat diperlukan oleh seluruh makhluk hidup. Air selalu berkaitan erat dengan keberadaan makhluk hidup biologis dan kehidupannya dalam alam sebagai tempat tumbuh dan berkembang biak. Dalam kehidupan sehari-hari,

manusia sangat tergantung pada air, dan kualitas kesehatan juga ditentukan oleh kualitas air yang dipergunakan untuk keperluan sehari-hari. Pemanfaatan air dapat diperuntukan sebagai sumber air minum, untuk keperluan rumah tangga, keperluan industri, memenuhi kebutuhan irigasi pertanian dan perkebunan, perikanan, sarana rekreasi dan lain-lain. (Situmorang, 2017).

2.3 Klasifikasi

Klasifikasi adalah sebuah teknik yang dilakukan dengan cara mengamati pada atribut dan kelakuan dari kelompok yang sudah didefinisikan. Teknik dengan cara memanipulasi data yang sudah ada dan sudah diklasifikasi kemudian menggunakan hasilnya untuk memberikan sejumlah aturan untuk melakukan klasifikasi pada data baru. Aturan-aturan yang telah dihasilkan kemudian digunakan untuk klasifikasi pada data-data baru. Teknik yang menggunakan kumpulan *record* dan pengujian yang telah diklasifikasi sebelumnya untuk menentukan kelas-kelas tambahan (Kusnawi, 2007).

Proses klasifikasi terbagi dari dua fase, yaitu fase learning dan fase *testing*. Pada fase *learning* dibentuk sebuah model perkiraan dengan menggunakan sebagian data yang telah diketahui kelas datanya. Sedangkan pada fase *testing* dilakukan pengujian untuk memperoleh tingkat akurasi pada model yang sudah terbentuk dengan menggunakan data lain.

2.4 Algoritme *Naïve Bayes*

Naïve Bayes merupakan metode pengklasifikasian suatu probabilitas dan statistik yang diperoleh Thomas Bayes seorang ilmuwan Inggris dengan cara melakukan prediksi peluang di masa depan berdasarkan pengalaman pada masa sebelumnya(Bustami, 2013). *Naïve Bayes* merupakan sebuah teknik prediksi probabilitas sederhana yang berdasarkan pada penerapan teorema *bayes* (aturan *bayes*) yang memiliki ketidakterkaitan antara suatu fitur dengan fitur lain dalam suatu data(Prasetyo, 2012). Formula *theorema bayes* dinyatakan dalam persamaan 2.1

$$P(H|E) = \frac{p(E|H)P(H)}{P(E)} \quad (2.1)$$

- H adalah hipotesis data X yang merupakan suatu kelas spesifik
- E adalah data dengan kelas yang belum diketahui
- P(H|E) adalah probabilitas hipotesis H berdasar *evidence*/bukti E (*posteriori probability*).
- P(H) adalah probabilitas hipotesis H (*prior probability*)
- P(E|H) adalah probabilitas *evidence* E berdasar kondisi hipotesis H.
- P(E) adalah probabilitas dari *evidence* E

2.5 Algoritme *K-Nearest Neighbor*

Prinsip kerja algoritme *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat dengan *k* tetangga (*neighbor*) terdekat dalam data *training* dengan data yang akan dievaluasi. Teknik mengelompokkan data baru dengan cara menghitung jarak data baru ke beberapa data/tetangga (*neighbor*) terdekat. Algoritme *K-Nearest Neighbor* merupakan *instead-based learning*, dimana data *training* disimpan sehingga klasifikasi untuk *record* baru yang belum diklasifikasi dapat ditemukan dengan membandingkan kemiripan yang paling banyak dalam data *training* (Kustiyahningsih, 2013).

Kelebihan dari algoritme *K-Nearest Neighbor* sendiri, yaitu:

1. Sederhana dalam penggunaannya.
2. Dapat menangani data *training* yang memiliki banyak *noise*
3. Efektif jika data *training* besar.

Kelemahan dari algoritme *K-Nearest Neighbor*, yaitu:

1. Algoritme *K-Nearest Neighbor* harus menentukan nilai parameter *K* (jumlah dari tetangga terdekat).
2. *Training* berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan.

Untuk menghitung jarak dalam *K-Nearest Neighbor* digunakan fungsi *Euclidean Distance* yang ditunjukkan pada Persamaan 2.3 (Kustiyahningsih, 2013):

$$euc = \sqrt{\sum_{i=1}^{n}(x_{2i} - x_{1i})^2} \quad (2.3)$$

Keterangan :

- X_2 = data latih
- X_1 = data uji
- i = variabel data
- n = dimensi data

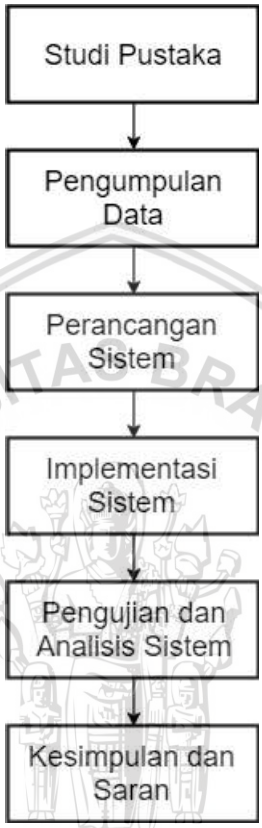
2.6 Pengujian Akurasi

Pengujian akurasi adalah suatu ukuran seberapa dekat hasil pengukuran terhadap angka sebenarnya. Akurasi dapat diperoleh dari persentase kebenaran, yaitu perbandingan antara jumlah data benar dengan keseluruhan data. Akurasi dinyatakan dengan rumus:

$$akurasi = \frac{total\ data\ benar}{total\ data}\ x\ 100\% \quad (2.4)$$

BAB 3 METODOLOGI

Pada bab ini berisikan alur metodologi penelitian dan yang akan digunakan dalam penelitian. Metodologi penelitian digunakan sebagai pedoman dalam pelaksanaan penelitian. Tahapan-tahapan metode penelitian dalam penelitian ini dapat ditunjukkan seperti pada Gambar 3.1



Gambar 3.1 Diagram metode tahapan penelitian

3.1 Studi Pustaka

Tahap ini dilakukan berdasarkan proses membaca dan pencarian informasi dari literatur, jurnal, skripsi, internet dan hasil penelitian serupa mengenai pemahaman pengertian air, pengertian air bersih, *K-Nearest Neighbor* dan *Naïve Bayes* sebagai algoritme yang digunakan, pengertian klasifikasi, serta persyaratan mengenai air bersih.

3.2 Pengumpulan Data

Penelitian ini menggunakan data tentang kualitas air bersih yang diperoleh dari PDAM Tirta Kencana Kabupaten Jombang di bawah pengawasan Dinas Kesehatan Kabupaten Jombang dan Balai Besar Teknik Kesehatan Lingkungan dan Pengendalian Penyakit Surabaya dalam jangka waktu tahun 2016 hingga 2017. Data ini menggunakan 5 parameter atau atribut, yaitu Bakteri *Coliform*, Bakteri *Escherichia coli*, Mangan (Mn), Total Padatan Terlarut (TDS) dan Khlorida (Cl).

3.3 Implementasi

Pada tahapan ini penerapan metode *K-Nearest Neighbor* dan *Naïve Bayes* dalam penelitian klasifikasi kualitas air bersih berdasarkan perancangan sistem. Di dalam implementasi ini, spesifikasi kebutuhan perangkat lunak maupun keras dalam pembuatan sistem klasifikasi kualitas air bersih berdasarkan metode *K-Nearest Neighbor* dan *Naïve Bayes* agar dapat berjalan dengan baik menggunakan sistem perangkat sebagai berikut :

a. Spesifikasi kebutuhan perangkat keras :
   - Laptop ASUS A456U Intel(R) Core(TM) i5-6200U CPU @2.30GHz 2.10 GHz
   - *Memory* RAM 8,00 GB 64-bit
   - *Hardisk* berkapasitas 1 TB
b. Spesifikasi kebutuhan perangkat lunak :
   - Windows 10 sebagai sistem operasi
   - Notepad++ dan Spyder sebagai *IDE*
   - Microsoft Office sebagai perhitungan manual, penempatan data dan sarana presentasi.

3.4 Lokasi Penelitian

Penelitian yang dilakukan oleh peneliti dilakukan di Laboratorium Riset Komputasi Cerdas, Fakultas Ilmu Komputer, Universitas Brawijaya Malang.

3.5 Jadwal Kegiatan

Rencana jadwal kegiatan penelitian ini dapat dilihat pada Tabel 3.1.

Tabel 3.1 Tabel Jadwal Kegiatan

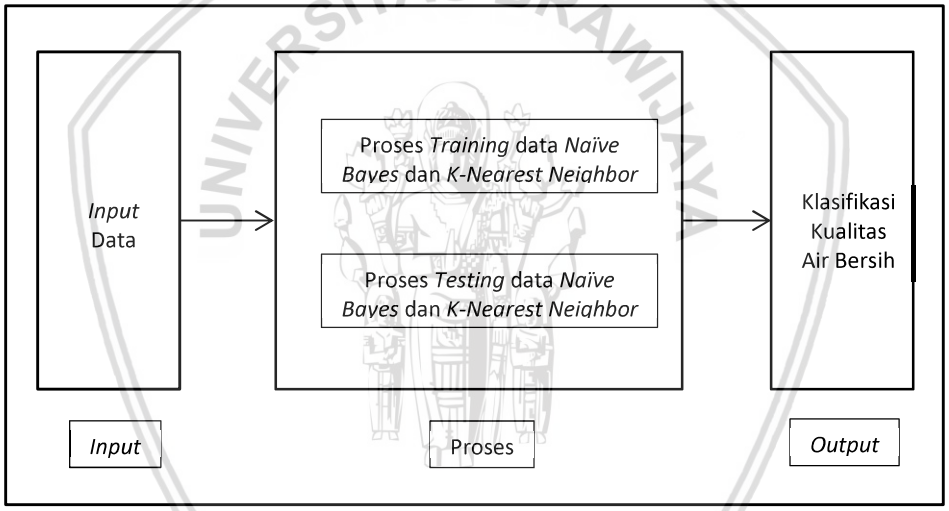
No	Kegiatan	Bulan			
		1	2	3	4
1	Studi Pustaka				
2	Pengumpulan Data				
3	Perancangan				
4	Implementasi				
5	Pengajuan dan Analisis				
6	Kesimpulan dan Saran				

BAB 4 PERANCANGAN DAN IMPLEMENTASI SISTEM

Pada bab ini kebutuhan perancangan yang akan digunakan untuk membuat aplikasi klasifikasi kualitas air bersih akan dibahas secara menyeluruh dan lengkap. Perancangan yang akan dibahas pada bab ini antara lain perancangan sistem, perancangan pengujian, perancangan antarmuka dan perancangan algoritme beserta manualisasi. Selain itu bab ini berisi tentang implementasi perangkat lunak dan perangkat kerasberdasarkan hasil analisis kebutuhan dan perancangan yang telah dilakukan.

4.1 Perancangan Sistem

Langkah-langkah penelitian yang dilakukan secara terstruktur mulai dari memasukkan data hingga memperoleh hasil. Langkah-langkah tersebut terdiri dari tiga proses, yaitu proses *input,* proses perhitungan dan proses *output.* Tiga proses tersebut akan dijelaskan pada Gambar 4.1.



Gambar 4.1 Diagram Perancangan Klasifikasi Kualitas Air Bersih

Keterangan Gambar 4.1 :

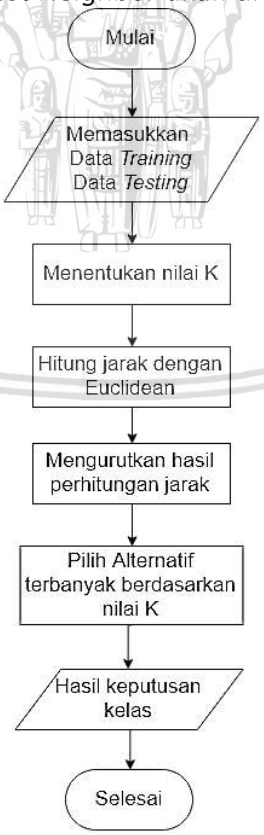
- Proses *Input*

  Masukan pada klasifikasi kualitas air bersih adalah atribut atau parameter air bersih yaitu Bakteri *Coliform*, Bakteri *Escherichia coli*, Mangan (Mn), Total Padatan Terlarut (TDS) dan Khlorida (Cl). Selain atribut tersebut terdapat nilai masukan berupa nilai *K.*

- Proses Perhitungan
  Untuk menentukan kelas kualitas air bersih pada penelitian menggunakan proses perhitungan menggunakan metode K-Nearest Neighbor dan Naïve Bayes. Terdapat dua langkah-langkah pada proses perhitungan yaitu menghitung data *training* menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes* dan menghitung data *testing* menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*.
- Proses *Output*
  Hasil yang diperoleh dari proses perhitungan yaitu klasifikasi untuk menentukan kualitas air bersih beserta dengan tingkat akurasinya.

4.1.1 Klasifikasi Menggunakan Metode *K-Nearest Neighbor*

Metode klasifikasi K-Nearest Neighbor memerlukan data *input* berupa data *training*, data *testing* dan nilai *K* untuk menunjukan jumlah tetangga terdekat (data *training* dengan data *testing*). Kemudian menggunakan rumus jarak *Euclidean* untuk menghitung jarak antara data *testing* dengan setiap data *training*. Setelah itu dilakukan pengurutan atau *sorting* data *training* berdasarkan jarak terkecil terhadap data *testing* hingga jarak terbesar antara data *training* terhadap data *testing*. Lalu mengambil data training terdekat sejumlah *K*. Dari *K* data *training* tersebut, dilihat mana kelas yang paling banyak muncul, maka kelas tersebut merupakan keputusan klasifikasi sistem. Proses perhitungan menggunakan metode *K-Nearest Neighbor* akan ditampilkan pada Gambar 4.2



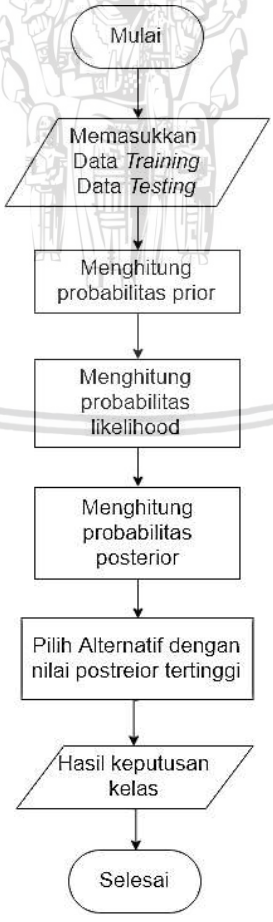
Gambar 4.2 Diagram alir klasifikasi K-Nearest Neighbor

4.1.2 Klasifikasi Menggunakan Metode *Naïve Bayes*

Pada metode *Naïve Bayes*, sistem melakukan proses klasifikasi pada data training. Langkah-langkah proses klasifikasi *Naïve Bayes* sebagai berikut :

1. Memasukan data pada tiap pa rameter atau atribut pada sistem untuk proses perhitungan untuk menentukan kategori atau kelas yang digunakan untuk proses perhitungan klasifikasi *Naïve Bayes*.
2. Melakukan perhitungan untuk mencari nilai *prior* dari masing-masing kelas. Kelas sendiri merupakan keputusan sistem yaitu kelas sesuai syarat atau batas dan kelas tidak sesuai syarat atau melewati batas yang ditetapkan.
3. Melakukan perhitungan untuk mencari nilai *likelihood* dari setiap kategori yang ada pada kelas sesuai syarat atau batas dan kelas tidak sesuai syarat atau melewati batas. Kategori tersebut meliputi kategori *coliform*, *escherichia coli*, mangan, tds dan khlorida.
4. Melakukan perhitungan untuk mencari nilai *posterior* dengan melukan proses perkalian semua nilai *prior* dan *likelihood* dari masing-masing kategori.

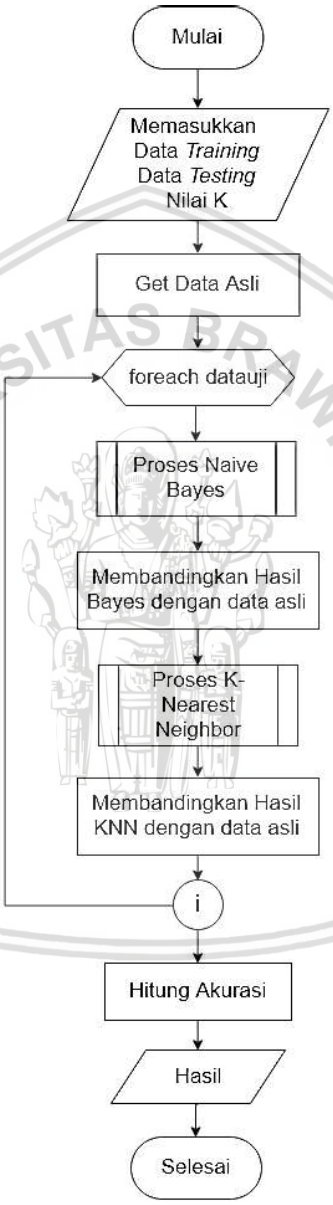
Kemudian akan didapatkan nilai peluang *posterior*. Rekomendasi penentuan klasifikasi kualitas air bersih menggunakan nilai peluang *posterior* terbesar dari masing-masing kelas. Diagram alir perhitungan menggunakan *Naive Bayes* dapat dilihat pada Gambar 4.3



Gambar 4.3 Diagram alir klasifikasi Naïve Bayes

4.1.3 Proses Perhitungan Semua Data

Proses hitung semua dapat dilihat pada Gambar 4.4. Pada diagram tersebut telah digambarkan bahwa proses dimulai ketika pengguna memasukkan jumlah data *training*, jumlah data *testing*, dan nilai k. Data *training* dan data *testing* tersebut dihitung menggunakan *K-Nearest Neighbor* dan *Naïve Bayes*. Selanjutnya dihitung akurasi yang dihasilkan dari kedua metode tersebut. Hasil yang ditampilkan berupa tabel kesesuaian antara kedua metode tersebut dengan data asli serta akurasi sistem.



Gambar 4.4 Diagram Alir Proses Perhitungan Semua Data

4.2 Perancangan Pengujian

Perancangan pengujian merupakan gambaran dari pengujian terhadap sistem klasifikasi kualitas air bersih yang menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*. Terdapat tiga model model pengujian yang terdapat pada perancangan pengujian penelitian ini, yaitu pengujian berdasarkan nilai *k* pada metode *K-Nearest Neighbor*, pengujian berdasarkan rasio perbandingan terhadap data *training* dan data *testing* dan pengujian terhadap variasi perbedaan jumlah data *training.*

4.2.1 Pengujian Berdasarkan Nilai *K* pada Metode *K-Nearest Neighbor*

Pada pengujian ini atribut nilai *K* dilihat untuk mengetahui terhadap metode *K-Nearest Neighbor.* Nilai *K* yang paling baik nantinya akan digunakan untuk perbandingan dengan metode *K-Nearest Neighbor* dengan metode *Naïve Bayes.* Dalam pengujian ditentukan nilai *K* = 1 ... 10 dimana setiap skenario pengujian *K* dilakukan sebanyak 5 kali. Dalam pencarian atribut *K* menggunakan data *testing* sebanyak 40 dan data *training* sejumlah 127. Contoh Tabel dari pengujian dapat dilihat seperti pada Tabel 4.1.

Tabel 4.1 Contoh Tabel Pengujian Berdasarkan Nilai Atribut K

K	Uji Coba Ke -	Akurasi (%)
1		
...		
10		

4.2.2 Pengujian Berdasarkan Rasio Perbandingan atau *Percentage Split*

Pada pengujian rasio perbandingan atau *percentage split* menggunakan data sejumlah 100% dari keseluruhan data atau sejumlah 167 data yang akan dibagi secara *random* atau acak berdasarkan perbandingan yang ditentukan yaitu menggunakan 90% data *training* dan 10% data *testing*, 80% data *training* dan 20% data *testing*, 70% data *training* dan 30% data *testing,* 60% data *training* dan 40% data *testing* dengan menggunakan nilai *K* terbaik untuk metode *K-Nearest Neighbor* dan metode *Naïve Bayes* berdasarkan variasi rasio *dataset* yang digunakan. Dari tiap skenario dilakukan 5 kali percobaan pengujian kemudian diambil rata-rata dari setiap skenario. Contoh Tabel dari pengujian berdasarkan rasio berbandingan dapat dilihat seperti pada Tabel 4.2.

Tabel 4.2 Contoh Tabel Pengujian Berdasarkan Rasio Perbandingan

Uji Coba Ke -	Data *Training*	Data *Testing*	Akurasi *K-Nearest Neighbor(%)*	Akurasi *Naïve Bayes(%)*

4.2.3 Pengujian Berdasarkan Jumlah Data *Training*

Pengujian berdasarkan variasi jumlah data *training* ini berbeda dengan pengujian sebelumnya yaitu *percentage split* dimana data yang dihgunakan 100% dari keseluruhan data. Pada pengujian ini tidak menggunakan data seluruhnya, hanya menggunakan beberapa data yang nantinya akan dibagi dalam jumlah data *training* yang berbeda. Pengujian ini dilakukan dengan jumlah data *training* yaitu 60 data, 80 data, 100 data, dan 120 data. Dan menggunakan jumlah data *testing* yang sama yaitu 40 data. Dari tiap skenario dilakukan 5 kali percobaan pengujian kemudian diambil rata-rata dari setiap skenario. Pengujian ini dilakukan untuk melihat komparasi tingkat akurasi dari metode *K-Nearest Neighbor* dan *Naïve Bayes* berdasarkan variasi jumlah data *training*. Contoh Tabel hasil pengujian berdasarkan variasi jumlah data *training* dapat ditunjukkan pada Tabel 4.3.

Tabel 4.3 Contoh Tabel Pengujian Berdasarkan Data Training

Uji Coba Ke -	Data *Training*	Data *Testing*	Akurasi *K-Nearest Neighbor(%)*	Akurasi *Naïve Bayes(%)*

4.3 Perancangan Algoritme

Perancangan algoritme merupakan gambaran dari cara kerja algoritme yang digunakan dalam penelitian, yaitu *K-Nearest Neighbor* dan *Naïve Bayes.* Pada perancangan algoritme ini akan dijelaskan bagaimana perhitungan manual atau manualisasi dari algoritme yang dilakukan untuk lebih memperjelas langkah-langkah dalam pembuatan sistem klasifikasi kualitas air bersih. Sebagai contoh kasus, terdapat 50 data training seperti pada Tabel 4.4.

Tabel 4.4 Contoh Kasus Data Training Perhitungan Manual

No.	*Coliform*	*E. Coli*	Mangan	TDS	Khlorida	Kelas
1	0	0	0	152	12,29	Sesuai Syarat
2	0	0	0	72	12,9	Sesuai Syarat
3	23	0	0,5	201	9,4	Tidak Sesuai Syarat
4	23	0	0	235	15,8	Tidak Sesuai Syarat
5	0	0	2	225	17,3	Sesuai Syarat
6	0	0	0	536	51,84	Tidak Sesuai Syarat
7	0	0	1,13	247	19,8	Sesuai Syarat
8	4,5	2	0	139	15	Tidak Sesuai Syarat
9	0	0	0	188	12,19	Sesuai Syarat
10	0	0	0	148	11,59	Sesuai Syarat
11	0	0	0,035	60	11,4	Sesuai Syarat
12	0,5	0	0,011	600	13,4	Tidak Sesuai Syarat
13	0	0	0,04	167	15,3	Sesuai Syarat
14	0	0	0,016	60	13,4	Sesuai Syarat
15	0	0	0,427	230	22,3	Sesuai Syarat
16	0	0	0	185	14,3	Sesuai Syarat

17	0	0	0	215	15,8	Sesuai Syarat
18	1	0	0,724	611	13,4	Tidak Sesuai Syarat
19	0	0	0	162	14,3	Sesuai Syarat
20	0	0	0	72	14,3	Sesuai Syarat
21	0	0	0	82	16	Sesuai Syarat
22	0	0	0	65	11,9	Sesuai Syarat
23	0	0	0	60	11,4	Sesuai Syarat
24	0	0	0	60	11	Sesuai Syarat
25	0	2,4	0	106	280	Tidak Sesuai Syarat
26	0	0	0	72	11,4	Sesuai Syarat
27	0	0	0	145	14,4	Sesuai Syarat
28	0	0	0	72	0	Sesuai Syarat
29	0	0	0	77	14,3	Sesuai Syarat
30	0	0	0,67	230	16,3	Sesuai Syarat
31	0	0	0	215	13,4	Sesuai Syarat
32	4,5	0	0	240	13,8	Tidak Sesuai Syarat
33	0	0	0	255	15,3	Sesuai Syarat
34	0	2	0,21	220	10,4	Tidak Sesuai Syarat
35	0	0	0	250	16,3	Sesuai Syarat
36	0	0	0	75	10,4	Sesuai Syarat
37	0	0	0	85	8,9	Sesuai Syarat
38	2	0	0,588	70	8,4	Tidak Sesuai Syarat
39	0	0	0	245	17,8	Sesuai Syarat
40	1	0	0	720	9,9	Tidak Sesuai Syarat
41	0	0	0	77	14,3	Sesuai Syarat
42	0	0	0,665	190	53,1	Sesuai Syarat
43	0	0	0,645	190	16,8	Sesuai Syarat
44	0	0	0	271	13,5	Sesuai Syarat
45	0	0	0	314	22,7	Sesuai Syarat
46	0	0	0,003	882	17,3	Tidak Sesuai Syarat
47	0	0	0	174	6,8	Sesuai Syarat
48	0	0	2,31	356	358	Tidak Sesuai Syarat
49	0	0	0	412	17,3	Sesuai Syarat
50	1,3	0	0,683	537	67,5	Tidak Sesuai Syarat

Diketahui data *testing* seperti pada tabel 4.5 dan parameter nilai k yang digunakan adalah 3.

Tabel 4.5 Contoh Kasus Dara Testing Perhitungan Manual

No.	*Coliform*	*E. Coli*	Mangan	TDS	Khlorida	Kelas
1	0	0	0	408	14,9	?
2	0	0	0	342	10	?
3	0	0	0	269	23,5	?
4	2	1,4	0	139	15	?
5	0	0	0,643	490	38,7	?

Sebelum dilakukan perhitungan, data *training* dan data *testing* diubah menjadi data numerik seperti format yang telah dilakukan sebagai berikut. Tabel 3.6 dan Tabel 3.7 merupakan Tabel data *training* dan data *testing* yang telah diubah menjadi data numerik.

Kelas = K6

- Tidak sesuai atau melebihi batas syarat bernilai 0
- Sesuai atau tidak melebihi batas syarat bernilai 1

4.3.1 Perhitungan dengan *K-Nearest Neighbor*

Perhitungan dilakukan dengan menghitung jarak antara data *testing* dengan setiap data *training*. Terdapat 50 data yang digunakan sebagai data *training*. Pada perhitungan jarak digunakan rumus jarak *Euclidean* seperti pada persamaan 2.2. Hasil dari perhitungan jarak *Euclidean* dapat dilihat pada Tabel 4.6.

Tabel 4.6 Hasil Perhitungan Jarak Euclidean

Data Testing 1			*Data Testing* 2		
No	Euc	Kelas	No	Euc	Kelas
1	256,0133045	1	1	190,0137998	1
2	336,0059523	1	2	270,0155736	1
3	208,3470662	0	3	142,8657062	0
4	174,5245255	0	4	109,5976277	0
5	183,0266647	1	5	117,2445734	1
6	133,2237351	0	6	198,4605392	0
7	161,0785116	1	7	95,51082085	1
8	269,0450892	0	8	203,1212692	0
9	220,0166905	1	9	154,015571	1
10	260,0210686	1	10	194,0065156	1
11	348,0176019	1	11	282,0034773	1
12	192,0065106	0	12	258,0228868	0
13	241,0003353	1	13	175,0802433	1
14	348,0032331	1	14	282,0204962	1
15	178,1542655	1	15	112,6741866	1
16	223,0008072	1	16	157,0588743	1
17	193,0020984	1	17	127,132372	1
18	203,0092958	0	18	269,0243189	0

19	246,0007317	1
20	336,0005357	1
21	326,0018558	1
22	343,0131193	1
23	348,0176001	1
24	348,0218528	1
25	401,8554093	0
26	336,0182287	1
27	263,0004753	1
28	336,3302098	1
29	331,0005438	1
30	178,0067664	1
31	193,0058289	1
32	168,0638569	0
33	153,0005229	1
34	188,0646009	0
35	158,0062024	1
36	333,030404	1
37	323,0557227	1
38	338,0689216	0
39	163,0257955	1
40	312,0416639	0
41	331,0005438	1
42	221,322575	1
43	218,0092338	1
44	137,0071531	1
45	94,32306187	1
46	474,0060759	0
47	234,1401503	1
48	347,025858	0
49	4,664761516	1
50	139,3194763	0

19	180,0513538	1
20	270,0342386	1
21	260,0692216	1
22	277,0065162	1
23	282,0034752	1
24	282,001773	1
25	358,6108755	0
26	270,0036296	1
27	197,0491309	1
28	270,1851217	1
29	265,0348845	1
30	112,1790484	1
31	127,0455037	1
32	102,1699075	0
33	87,16128728	1
34	122,0172287	0
35	92,21545424	1
36	267,0002996	1
37	257,0023541	1
38	272,0126941	0
39	97,31310292	1
40	378,001336	0
41	265,0348845	1
42	157,993836	1
43	152,1533964	1
44	71,08621526	1
45	30,74556879	1
46	540,0493403	0
47	168,0304734	1
48	348,2891559	0
49	70,37961353	1
50	203,306189	0

Data Testing 3		
No	Euc	Kelas
1	117,5357992	1
2	197,2849716	1
3	73,15777471	0
4	41,76469801	0
5	44,47965827	1
6	268,4998242	0

Data Testing 4		
No	Euc	Kelas
1	13,50200356	1
2	67,07734342	1
3	65,71582762	0
4	98,28326409	0
5	86,08861713	1
6	398,7131119	0

7	22,33756701	1
8	130,3706255	0
9	81,785794	1
10	121,5847363	1
11	209,3499731	1
12	331,1544355	0
13	102,3290848	1
14	209,2439014	1
15	39,02079355	1
16	84,50230766	1
17	54,54621893	1
18	342,1513323	0
19	107,3947857	1
20	197,2147053	1
21	187,1503406	1
22	204,3295378	1
23	209,3499701	1
24	209,3734701	1
25	303,9194137	0
26	197,3712492	1
27	124,3334629	1
28	198,3966986	1
29	192,2202903	1
30	39,66470597	1
31	54,93641779	1
32	30,90857486	0
33	16,22467257	1
34	50,76075354	0
35	20,31846451	1
36	194,4417908	1
37	184,5783303	1
38	199,5829545	0
39	24,66759007	1
40	451,206117	0
41	192,2202903	1
42	84,36588306	1
43	79,28622847	1
44	10,19803903	1
45	45,00711055	1
46	613,0313532	0
47	96,45667421	1

7	108,14008	1
8	2,570992026	0
9	49,14118537	1
10	9,929154043	1
11	79,11966396	1
12	461,0073428	0
13	28,10785655	1
14	79,05390728	1
15	91,32596744	1
16	46,07005535	1
17	76,04340866	1
18	472,0064027	0
19	23,13979257	1
20	67,04811705	1
21	57,06101997	1
22	74,10512803	1
23	79,11965622	1
24	79,13886529	1
25	267,0561739	0
26	67,14104557	1
27	6,505382387	1
28	68,70196504	1
29	62,05199433	1
30	91,04448858	1
31	76,05603198	1
32	101,047761	0
33	116,0260747	1
34	81,15764967	0
35	111,0344541	1
36	64,21152545	1
37	54,39825365	1
38	69,33156384	0
39	106,0650744	1
40	581,0249306	0
41	62,05199433	1
42	63,71037769	1
43	51,09418778	1
44	132,0310948	1
45	175,1863294	1
46	743,0075706	0
47	36,0305426	1

48	345,6364942	0
49	143,1343425	1
50	271,5918933	0

48	405,8932078	0
49	273,0206036	1
50	401,4513252	0

Data Testing 5		
No	Euc	Kelas
1	339,0308268	1
2	418,7959568	1
3	291,3906492	0
4	257,0580157	0
5	265,8661345	1
6	47,84425827	0
7	243,7343783	1
8	351,8342699	0
9	303,1619923	1
10	343,0734113	1
11	430,8661737	1
12	112,8748839	0
13	323,8470682	1
14	430,744104	1
15	260,5168069	1
16	305,975119	1
17	275,9525746	1
18	123,6207772	0
19	328,9069374	1
20	418,7120412	1
21	408,6315008	1
22	425,8446353	1
23	430,8662245	1
24	430,8917537	1
25	453,5282389	0
26	418,8910401	1
27	345,8553216	1
28	419,788165	1
29	413,7206466	1
30	260,9631406	1
31	276,1620963	1
32	251,2780799	0
33	236,163023	1
34	271,4867906	0
35	241,0439243	1

36	415,9643055	1
37	406,0953748	1
38	421,096299	0
39	245,8906738	1
40	231,7991662	0
41	413,7206466	1
42	300,345402	1
43	300,7982879	1
44	220,446033	1
45	176,726946	1
46	392,5842198	0
47	317,6067119	1
48	346,2820655	0
49	80,88493957	1
50	55,1373884	0

Setelah didapatkan hasil dari perhitungan jarak *Euclidean*, selanjutnya dilakukan pengurutan atau sorting dari jarak *Euclidean* terendah. Hasil pengurutan dapat dilihat pada tabel 4.7.

Tabel 4.7 Hasil pengurutan jarak Euclidean

Data Testing 1			*Data Testing* 2		
No	Euc	Kelas	No	Euc	Kelas
49	4,664761516	1	45	30,74556879	1
45	94,32306187	1	49	70,37961353	1
6	133,2237351	0	44	71,08621526	1
44	137,0071531	1	33	87,16128728	1
50	139,3194763	0	35	92,21545424	1
33	153,0005229	1	7	95,51082085	1
35	158,0062024	1	39	97,31310292	1
7	161,0785116	1	32	102,1699075	0
39	163,0257955	1	4	109,5976277	0
32	168,0638569	0	30	112,1790484	1
4	174,5245255	0	15	112,6741866	1
30	178,0067664	1	5	117,2445734	1
15	178,1542655	1	34	122,0172287	0
5	183,0266647	1	31	127,0455037	1
34	188,0646009	0	17	127,132372	1
12	192,0065106	0	3	142,8657062	0
17	193,0020984	1	43	152,1533964	1
31	193,0058289	1	9	154,015571	1
18	203,0092958	0	16	157,0588743	1

3	208,3470662	0
43	218,0092338	1
9	220,0166905	1
42	221,322575	1
16	223,0008072	1
47	234,1401503	1
13	241,0003353	1
19	246,0007317	1
1	256,0133045	1
10	260,0210686	1
27	263,0004753	1
8	269,0450892	0
40	312,0416639	0
37	323,0557227	1
21	326,0018558	1
29	331,0005438	1
41	331,0005438	1
36	333,030404	1
20	336,0005357	1
2	336,0059523	1
26	336,0182287	1
28	336,3302098	1
38	338,0689216	0
22	343,0131193	1
48	347,025858	0
14	348,0032331	1
23	348,0176001	1
11	348,0176019	1
24	348,0218528	1
25	401,8554093	0
46	474,0060759	0

42	157,993836	1
47	168,0304734	1
13	175,0802433	1
19	180,0513538	1
1	190,0137998	1
10	194,0065156	1
27	197,0491309	1
6	198,4605392	0
8	203,1212692	0
50	203,306189	0
37	257,0023541	1
12	258,0228868	0
21	260,0692216	1
29	265,0348845	1
41	265,0348845	1
36	267,0002996	1
18	269,0243189	0
26	270,0036296	1
2	270,0155736	1
20	270,0342386	1
28	270,1851217	1
38	272,0126941	0
22	277,0065162	1
24	282,001773	1
23	282,0034752	1
11	282,0034773	1
14	282,0204962	1
48	348,2891559	0
25	358,6108755	0
40	378,001336	0
46	540,0493403	0

Data Testing 3		
No	Euc	Kelas
44	10,19803903	1
33	16,22467257	1
35	20,31846451	1
7	22,33756701	1
39	24,66759007	1
32	30,90857486	0
15	39,02079355	1
30	39,66470597	1
4	41,76469801	0

Data Testing 4		
No	Euc	Kelas
8	2,570992026	0
27	6,505382387	1
10	9,929154043	1
1	13,50200356	1
19	23,13979257	1
13	28,10785655	1
47	36,0305426	1
16	46,07005535	1
9	49,14118537	1

5	44,47965827	1
45	45,00711055	1
34	50,76075354	0
17	54,54621893	1
31	54,93641779	1
3	73,15777471	0
43	79,28622847	1
9	81,785794	1
42	84,36588306	1
16	84,50230766	1
47	96,45667421	1
13	102,3290848	1
19	107,3947857	1
1	117,5357992	1
10	121,5847363	1
27	124,3334629	1
8	130,3706255	0
49	143,1343425	1
37	184,5783303	1
21	187,1503406	1
29	192,2202903	1
41	192,2202903	1
36	194,4417908	1
20	197,2147053	1
2	197,2849716	1
26	197,3712492	1
28	198,3966986	1
38	199,5829545	0
22	204,3295378	1
14	209,2439014	1
23	209,3499701	1
11	209,3499731	1
24	209,3734701	1
6	268,4998242	0
50	271,5918933	0
25	303,9194137	0
12	331,1544355	0
18	342,1513323	0
48	345,6364942	0
40	451,206117	0
46	613,0313532	0

43	51,09418778	1
37	54,39825365	1
21	57,06101997	1
29	62,05199433	1
41	62,05199433	1
42	63,71037769	1
36	64,21152545	1
3	65,71582762	0
20	67,04811705	1
2	67,07734342	1
26	67,14104557	1
28	68,70196504	1
38	69,33156384	0
22	74,10512803	1
17	76,04340866	1
31	76,05603198	1
14	79,05390728	1
23	79,11965622	1
11	79,11966396	1
24	79,13886529	1
34	81,15764967	0
5	86,08861713	1
30	91,04448858	1
15	91,32596744	1
4	98,28326409	0
32	101,047761	0
39	106,0650744	1
7	108,14008	1
35	111,0344541	1
33	116,0260747	1
44	132,0310948	1
45	175,1863294	1
25	267,0561739	0
49	273,0206036	1
6	398,7131119	0
50	401,4513252	0
48	405,8932078	0
12	461,0073428	0
18	472,0064027	0
40	581,0249306	0
46	743,0075706	0

Data Testing 5		
No	Euc	Kelas
6	47,84425827	0
50	55,1373884	0
49	80,88493957	1
12	112,8748839	0
18	123,6207772	0
45	176,726946	1
44	220,446033	1
40	231,7991662	0
33	236,163023	1
35	241,0439243	1
7	243,7343783	1
39	245,8906738	1
32	251,2780799	0
4	257,0580157	0
15	260,5168069	1
30	260,9631406	1
5	265,8661345	1
34	271,4867906	0
17	275,9525746	1
31	276,1620963	1
3	291,3906492	0
42	300,345402	1
43	300,7982879	1
9	303,1619923	1
16	305,975119	1
47	317,6067119	1
13	323,8470682	1
19	328,9069374	1
1	339,0308268	1
10	343,0734113	1
27	345,8553216	1
48	346,2820655	0
8	351,8342699	0
46	392,5842198	0
37	406,0953748	1
21	408,6315008	1
29	413,7206466	1
41	413,7206466	1
36	415,9643055	1
20	418,7120412	1
2	418,7959568	1

26	418,8910401	1
28	419,788165	1
38	421,096299	0
22	425,8446353	1
14	430,744104	1
11	430,8661737	1
23	430,8662245	1
24	430,8917537	1
25	453,5282389	0

Kemudian hasil keputusan diambil berdasarkan k tetangga terdekat, seperti terlihat pada tabel 4.8.

Tabel 4.8 Hasil keputusan berdasarkan nilai K

Data	Hasil Nilai K	Keputusan
1	(1,1,0)	Sesuai Syarat
2	(1,1,1)	Sesuai Syarat
3	(1,1,1)	Sesuai Syarat
4	(0,1,1)	Sesuai Syarat
5	(0,0,1)	Tidak Sesuai Syarat

4.3.2 Perhitungan dengan *Naïve Bayes*

Perhitungan dengan *Naïve Bayes* dilakukan dengan menentukan nilai probabilitas prior terlebih dahulu berdasarkan data training. Hasil perhitungan probabilitas prior dapat dilihat pada Tabel 4.9.

Tabel 4.9 Perhitungan Proabilitas Prior

No	Probabilitas	Hasil	
1	P(K-06 = 0)	14/50	1
2	P(K-06 = 1)	36/50	2

Setelah didapatkan hasil probabilitas *prior*, langkah selanjutnya adalah perhitungan probabilitas *likelihood*. Probabilitas *likelihood* pada penelitian ini dibagi menjadi dua bagian, yaitu untuk kelas Tidak Sesuai Syarat dan kelas Sesuai Syarat. Hasil perhitungan probabilitas *likelihood* dapat dilihat pada Tabel 4.10 dan Tabel 4.11.

Tabel 4.10 Perhitungan Probabilitas Likelihood Tidak Sesuai

No	Probabilitas	Hasil	
1	P(K-01=0 | K-06=0)	5	0,357142857
2	P(K-01=1 | K-06=0)	9	0,642857143
3	P(K-02=0 | K-06=0)	3	0,214285714
4	P(K-02=1 | K-06=0)	11	0,785714286
5	P(K-03=0 | K-06=0)	5	0,357142857

6	P(K-03=1 | K-06=0)	9	0,642857143
7	P(K-04=0 | K-06=0)	6	0,428571429
8	P(K-04=1 | K-06=0)	8	0,571428571
9	P(K-05=0 | K-06=0)	2	0,142857143
10	P(K-05=1 | K-06=0)	12	0,857142857

Tabel 4.11 Perhitungan Probabilitas Likelihood Sesuai

No	Probabilitas	Hasil	
1	P(K-01=0 | K-06=1)	0	0
2	P(K-01=1 | K-06=1)	36	1
3	P(K-02=0 | K-06=1)	0	0
4	P(K-02=1 | K-06=1)	36	1
5	P(K-03=0 | K-06=1)	6	0,166666667
6	P(K-03=1 | K-06=1)	30	0,833333333
7	P(K-04=0 | K-06=1)	0	0
8	P(K-04=1 | K-06=1)	36	1
9	P(K-05=0 | K-06=1)	0	0
10	P(K-05=1 | K-06=1)	36	1

Setelah didapatkan hasil dari perhitungan probabilitas *likelihood*, langkah selanjutnya atau langkah yang terakhir adalah perhitungan probabilitas *posterior*. Perhitungan probabilitas *posterior* sebagai berikut :

Data *Testing* 1 :

- P(K-06=0 | E) = P(K-06=0) * P(K-01=1|K-06=0) * P(K-02=1|K-06=0) * P(K-03=1|K-06=0) * P(K-04=1|K-06=0) * P(K-05=1|K-06=0) = **0,044531445**
- P(K-06=1 | E) = P(K-06=1) * P(K-01=1|K-06=1) * P(K-02=1|K-06=1) * P(K-03=1|K-06=1) * P(K-04=1|K-06=1) * P(K-05=1|K-06=1) = **0,6**

Data *Testing* 2 :

- P(K-06=0 | E) = P(K-06=0) * P(K-01=1|K-06=0) * P(K-02=1|K-06=0) * P(K-03=1|K-06=0) * P(K-04=1|K-06=0) * P(K-05=1|K-06=0) = **0,044531445**
- P(K-06=1 | E) = P(K-06=1) * P(K-01=0|K-06=1) * P(K-02=0|K-06=1) * P(K-03=1|K-06=1) * P(K-04=0|K-06=1) * P(K-05=0|K-06=1) = **0,6**

Data *Testing* 3 :

- P(K-06=0 | E) = P(K-06=0) * P(K-01=1|K-06=0) * P(K-02=1|K-06=0) * P(K-03=1|K-06=0) * P(K-04=1|K-06=0) * P(K-05=1|K-06=0) = **0,044531445**

- P(K-06=1 | E) = P(K-06=1) * P(K-01=0|K-06=1) * P(K-02=0|K-06=1) * P(K-03=1|K-06=1) * P(K-04=0|K-06=1) * P(K-05=0|K-06=1) = **0,66**

Data *Testing* 4 :

- P(K-06=0 | E) = P(K-06=0) * P(K-01=0|K-06=0) * P(K-02=0|K-06=0) * P(K-03=1|K-06=0) * P(K-04=1|K-06=0) * P(K-05=1|K-06=0) = **0,006747189**
- P(K-06=1 | E) = P(K-06=1) * P(K-01=0|K-06=1) * P(K-02=0|K-06=1) * P(K-03=1|K-06=1) * P(K-04=1|K-06=1) * P(K-05=1|K-06=1) = **0**

Data *Testing* 5 :

- P(K-06=0 | E) = P(K-06=0) * P(K-01=1|K-06=0) * P(K-02=1|K-06=0) * P(K-03=0|K-06=0) * P(K-04=1|K-06=0) * P(K-05=1|K-06=0) = **0,024739692**
- P(K-06=1 | E) = P(K-06=1) * P(K-01=1|K-06=1) * P(K-02=1|K-06=1) * P(K-03=0|K-06=1) * P(K-04=1|K-06=1) * P(K-05=1|K-06=1) = **0,12**

Setelah dilakukan perhitungan probabilitas *posterior* seperti diatas, hasil keputusan klasifikasi data testing dapat dilihat seperti pada Tabel 4.12.

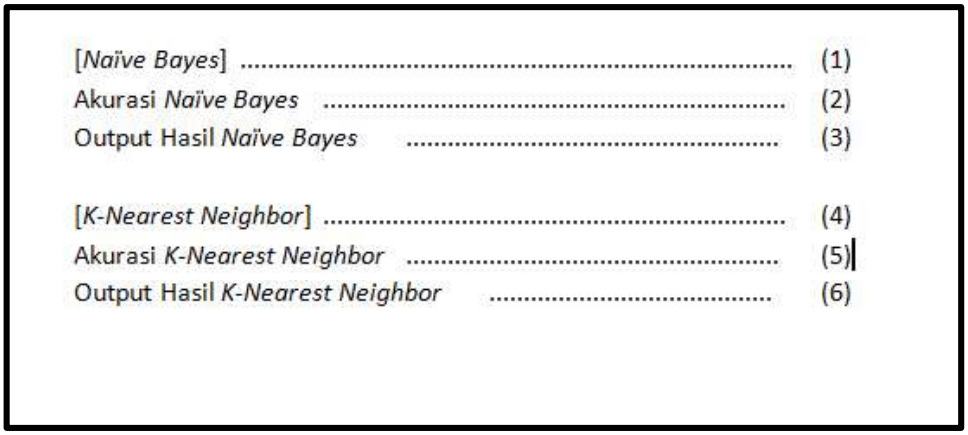
Tabel 4.12 Hasil Perhitungan Probabilitas Posterior

Data	Hasil Nilai K		Keputusan
	K-06 = 0	K-06 = 1	
1	0,044531445	0,6	Sesuai Syarat
2	0,044531445	0,6	Sesuai Syarat
3	0,044531445	0,6	Sesuai Syarat
4	0,006747189	0	Tidak Sesuai Syarat
5	0,024739692	0,12	Sesuai Syarat

4.4 Perancangan Antarmuka

Antarmuka pengguna atau yang sering disebut dengan *interface* merupakan tampilan antarmuka aplikasi program dengan pengguna. Hal ini bertujuan untuk membuat pengguna dalam menjalankan aplikasi program yang telah dirancang dengan mudah dan nyaman.

Perancangan ini memiliki tujuan untuk menampilkan dan memberi konsep gambaran hasil dari sistem dan implementasi metode *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi kualitas air bersih. Pada perancangan antarmuka penelitian ini berupa *Command Line Interface (CLI)* karena tidak menggunakan *Graphical User Interface (GUI)*. Perancangan antarmuka hasil dapat dilihat pada Gambar 4.5 berikut ini



Gambar 4.5 Perancangan Tampilan Antarmuka Hasil

Keterangan :

1. Nama metode yaitu *Naïve Bayes*
2. Nilai akurasi yang didapatkan metode *Naïve Bayes*
3. Hasil perhitungan dengan metode *Naïve Bayes*
4. Nama metode yaitu *K-Nearest Neighbor*
5. Nilai akurasi yang didapatkan metode *K-Nearest Neighbor*
6. Hasil perhitungan dengan metode *K-Nearest Neighbor*

4.5 Implementasi Sistem

Pada implementasi sistem akan dibahas tentang spesifikasi-spesifikasi yang digunakan dalam melakukan implementasi sistem. Spesifikasi sistem terbagi menjadi dua bagian yaitu spesifikasi perangkat lunak atau *software* dan spesifikasi perangkan keras atau *hardware.*

4.5.1 Spesifikasi Perangkat Lunak

Spesifikasi perangkat lunak atau *software* yang digunakan pada implementasi komparasi metode *K-Nearest Neighbor* dan *Naïve Bayes* dalam klasifikasi kualitas air bersih dapat dilihat secara terperinci pada Tabel 4.13.

Tabel 4.13 Spesifikasi Perangkat Lunak atau Software

Nama Komponen	Spesifikasi
Sistem operasi	Windows 10 Pro
Basis data	Microsoft Excel 2010
Tools dokumentasi	Microsoft Office 2010
Tools diagram	Draw.io
Bahasa pemprograman	Python 2.7
Tools pemprograman/*IDE*	Anaconda Spyder 3

4.5.2 Spesifikasi Perangkat Keras

Spesifikasi perangkat keras atau *hardware* yang digunakan pada implementasi komparasi metode *K-Nearest Neighbor* dan *Naïve Bayes* dalam klasifikasi kualitas air bersih dapat dilihat secara terperinci pada Tabel 4.14.

Tabel 4.14 Spesifikasi Peangkat Keras atau Hardware

Nama Komponen	Spesifikasi
Prosesor	Intel(R) Core(TM) i5-6200U @2.30GHz 2.10 GHz
Memori (RAM)	8 GB
Hard disk	1 TB
Kartu grafis	Intel(R) HD Graphics 520 dan NVIDIA GeForce 930MX
Monitor	14.1”

4.6 Implementasi Algoritme

Implementasi algoritme mengacu pada sistem model klasifikasi yang telah dijelaskan pada bab sebelumnya. Sub bab ini berisikan tentang implementasi algoritme berupa *source code* atau kode program dari sistem klasifikasi kualitas air bersih yang menggunakan algoritme *K-Nearest Neighbor* dan *Naïve Bayes.* Kode program implementasi algoritme *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi kualitas air bersih dapat dilihat pada tabel 4.15.

Tabel 4.15 Kode Program Implementasi Algoritme

Baris	Kode
1	`import pandas as pd`
2	`from sklearn.model_selection import train_test_split`
3	`from sklearn.metrics import accuracy_score`
4	`import math`
5	`#Bayes`
6	`df = pd.read_excel('C:\Users\Owner\Desktop\Bayes.xlsx')`
7	
8	`coliform = df['Coliform'].tolist()`
9	`ecoli = df['E.Coli'].tolist()`
10	`mangan = df['Mangan'].tolist()`
11	`tds = df['TDS'].tolist()`
12	`khlorida = df['Khlorida'].tolist()`
13	`output_target = df['Kelas'].tolist()`
14	
15	`fitur = [list(l) for l in zip(coliform, ecoli, mangan, tds,`
16	`khlorida)]`
17	
18	`fiturTrain, fiturTest, outputTrain, outputTest =`
19	`train_test_split(fitur,`
20	
21	`output_target,`
22	
23	`test_size = 0.3)`
24	
25	`coliform00 = 0`
26	`coliform01 = 0`

```
coliform11 = 0
coliform10 = 0

ecoli00 = 0
ecoli01 = 0
ecoli10 = 0
ecoli11 = 0

mangan00 = 0
mangan01 = 0
mangan10 = 0
mangan11 = 0

tds00 = 0
tds01 = 0
tds10 = 0
tds11 = 0

khlorida00 = 0
khlorida01 = 0
khlorida10 = 0
khlorida11 = 0

for i in zip(coliform, outputTrain):
    if i == (0, 0):
        coliform00 += 1
    elif i == (0, 1):
        coliform01 +=1
    elif i == (1, 0):
        coliform10 +=1
    elif i == (1, 1):
        coliform11 +=1

for i in zip(ecoli, outputTrain):
    if i == (0, 0):
        ecoli00 += 1
    elif i == (0, 1):
        ecoli01 +=1
    elif i == (1, 0):
        ecoli10 +=1
    elif i == (1, 1):
        ecoli11 +=1

for i in zip(mangan, outputTrain):
    if i == (0, 0):
        mangan00 += 1
```

```
    elif i == (0, 1):
        mangan01 +=1
    elif i == (1, 0):
        mangan10 +=1
    elif i == (1, 1):
        mangan11 +=1

for i in zip(tds, outputTrain):
    if i == (0, 0):
        tds00 += 1
    elif i == (0, 1):
        tds01 +=1
    elif i == (1, 0):
        tds10 +=1
    elif i == (1, 1):
        tds11 +=1

for i in zip(khlorida, outputTrain):
    if i == (0, 0):
        khlorida00 += 1
    elif i == (0, 1):
        khlorida01 +=1
    elif i == (1, 0):
        khlorida10 +=1
    elif i == (1, 1):
        khlorida11 +=1

kelas0 = 0
kelas1 = 0

for i in outputTrain:
    if i == 0:
        kelas0 += 1
    elif i == 1:
        kelas1 += 1

p_kelas_0 = float(kelas0) / len(outputTrain)
p_kelas_1 = float(kelas1) / len(outputTrain)

p_coliform_00 = float(coliform00) / kelas0
p_coliform_01 = float(coliform01) / kelas1
p_coliform_10 = float(coliform10) / kelas0
p_coliform_11 = float(coliform11) / kelas1

p_ecoli_00 = float(ecoli00) / kelas0
p_ecoli_01 = float(ecoli01) / kelas1
```

```
119 p_ecoli_10 = float(ecoli10) / kelas0
120 p_ecoli_11 = float(ecoli11) / kelas1
121
122 p_mangan_00 = float(mangan00) / kelas0
123 p_mangan_01 = float(mangan01) / kelas1
124 p_mangan_10 = float(mangan10) / kelas0
125 p_mangan_11 = float(mangan11) / kelas1
126
127 p_tds_00 = float(tds00) / kelas0
128 p_tds_01 = float(tds01) / kelas1
129 p_tds_10 = float(tds10) / kelas0
130 p_tds_11 = float(tds11) / kelas1
131
132 p_khlorida_00 = float(khlorida00) / kelas0
133 p_khlorida_01 = float(khlorida01) / kelas1
134 p_khlorida_10 = float(khlorida10) / kelas0
135 p_khlorida_11 = float(khlorida11) / kelas1
136
137 p_coliform_0_temp = 0
138 p_coliform_1_temp = 0
139 p_ecoli_0_temp = 0
140 p_ecoli_1_temp = 0
141 p_mangan_0_temp = 0
142 p_mangan_1_temp = 0
143 p_tds_0_temp = 0
144 p_tds_1_temp = 0
145 p_khlorida_0_temp = 0
146 p_khlorida_1_temp = 0
147
148 pred = []
149
150 for i in range(len(fiturTest)):
151     coliform_var = fiturTest[i][0]
152     ecoli_var = fiturTest[i][1]
153     mangan_var = fiturTest[i][2]
154     tds_var = fiturTest[i][3]
155     khlorida_var = fiturTest[i][4]
156
157     if(coliform_var == 0):
158         p_coliform_0_temp = p_coliform_00
159         p_coliform_1_temp = p_coliform_01
160     elif(coliform_var == 1):
161         p_coliform_0_temp = p_coliform_10
162         p_coliform_1_temp = p_coliform_11
163
164     if(ecoli_var == 0):
```

```
165             p_ecoli_0_temp = p_ecoli_00
166             p_ecoli_1_temp = p_ecoli_01
167         elif(ecoli_var == 1):
168             p_ecoli_0_temp = p_ecoli_10
169             p_ecoli_1_temp = p_ecoli_11
170 
171         if(mangan_var == 0):
172             p_mangan_0_temp = p_mangan_00
173             p_mangan_1_temp = p_mangan_01
174         elif(mangan_var == 1):
175             p_mangan_0_temp = p_mangan_10
176             p_mangan_1_temp = p_mangan_11
177 
178         if(tds_var == 0):
179             p_tds_0_temp = p_tds_00
180             p_tds_1_temp = p_tds_01
181         elif(tds_var == 1):
182             p_tds_0_temp = p_tds_10
183             p_tds_1_temp = p_tds_11
184 
185         if(khlorida_var == 0):
186             p_khlorida_0_temp = p_khlorida_00
187             p_khlorida_1_temp = p_khlorida_01
188         elif(khlorida_var == 1):
189             p_khlorida_0_temp = p_khlorida_10
190             p_khlorida_1_temp = p_khlorida_11
191 
192         test_0 = p_coliform_0_temp * p_ecoli_0_temp *
193 p_mangan_0_temp * p_tds_0_temp * p_khlorida_0_temp *
194 p_kelas_0
195         test_1 = p_coliform_1_temp * p_ecoli_1_temp *
196 p_mangan_1_temp * p_tds_1_temp * p_khlorida_1_temp *
197 p_kelas_1
198 
199         if(test_0 > test_1):
200             pred.append(0)
201         else:
202             pred.append(1)
203 
204 print ""
205 print "[Naïve Bayes]"
206 print "Accuracy with Naïve Bayes:
207 "+str(accuracy_score(outputTest, pred) * 100)+"%"
208 print zip (outputTest, pred)
209 
210 #KNN FIX
```

```
df = pd.read_excel('C:\Users\Owner\Desktop\knn.xlsx')

coliform = df['Coliform'].tolist()
ecoli = df['E.Coli'].tolist()
mangan = df['Mangan'].tolist()
tds = df['TDS'].tolist()
khlorida = df['Khlorida'].tolist()
output_target = df['Kelas'].tolist()

fitur = [list(l) for l in zip(coliform, ecoli, mangan, tds,
khlorida)]

fiturTrain, fiturTest, outputTrain, outputTest =
train_test_split(fitur,

output_target,

test_size = 0.3)

n_neighbors = 3
euclidean = 0
euclidean_list_complete = []

for i in range(len(fiturTest)):
    euclidean_list = []
    for j in range(len(fiturTrain)):
        for k in range(len(fiturTrain[0])):
            euclidean += pow((fiturTest[i][k]-
fiturTrain[j][k]),2)
        euclidean = math.sqrt(euclidean)
        euclidean_list.append((euclidean, outputTrain[j]))
        euclidean = 0
        euclidean_list.sort()

euclidean_list_complete.append(euclidean_list[:n_neighbors])
    del euclidean_list

pre_vote = []
for i in range(len(euclidean_list_complete)):
    temp = []
    for j in range(len(euclidean_list_complete[0])):
        temp.append(euclidean_list_complete[i][j][1])
    pre_vote.append(temp)
    del temp

pred = []
```

```
257 for i in range(len(pre_vote)):
258     count0 = pre_vote[i].count(0)
259     count1 = pre_vote[i].count(1)
260     if count0 > count1:
261         pred.append(0)
262     else:
263         pred.append(1)
264
265 print ""
266 print "[K-Nearest Neighbor]"
267 print "Accuracy with K-NN: "+str(accuracy_score(outputTest,
268 pred) * 100)+"%"
269 print zip (outputTest, pred)
270
```

Penjelasan kode program :

Baris 1 : mengimport library untuk memanipulasi data

Baris 2 : mengimport fungsi untuk membagi dataset untuk data *training*

Baris 3 : mengimport fungsi untuk menghitung tingkat akurasi

Baris 6 : memanggil data dari *database*

Baris 8 - 13 : deklarasi variabel-variabel dari *database*

Baris 15 - 23 : deklarasi variabel

Baris 25 - 48 : inisialisasi nilai variabel

Baris 50 - 90 : fungsi perulangan untuk menghitung jumlah probabilitas tiap variabel

Baris 112 - 115 : proses peritungan probabilitas *prior*

Baris 137 - 146 : deklarasi variabel *temp*

Baris 150 - 190 : proses perhitungan probabilitas *likelihood*

Baris 192 - 197 : proses perhitungan probabilitas *posterior*

Baris 199 - 202 : proses menentukan kelas *output* klasifikasi

Baris 204 - 208 : menampilkan hasil perhitungan klasifikasi dengan nilai akurasi yang didapatkan

Baris 211 : memanggil data dari *database*

Baris 213 - 218 : deklarasi variabel-variabel dari *database*

Baris 220 - 228 : deklarasi variabel

Baris 230 - 232 : deklarasi variabel n_neighbors untuk menentukan nilai *K* dan variabel euclidean

Baris 234 - 246 : proses perhitungan jarak *euclidean* data *training* dengan data *testing*

Baris 249 - 254 : proses pengurutan data berdasarkan jarak terpendek euclidean

Baris 256 - 263 : menampilkan hasil perhitungan klasifikasi dengan nilai akurasi yang didapatkan

4.7 Implementasi Tampilan Antarmuka

Implementasi tampilan antarmuka merupakan gambaran dari sistem yang telah dibuat berdasarkan perancangan tampilan antarmuka yang telah dirancang pada bab sebelumnya. Tampilan antarmuka pada sistem dapat dilihat pada Gambar 4.6

```
In [1]: runfile('C:/Users/Owner/Desktop/knn_bayes_new.py', wdir='C:/Users/
Owner/Desktop')

[Naïve Bayes]
Accuracy with Naïve Bayes: 74.5098039216%
[(1L, 1), (0L, 0), (1L, 1), (0L, 1), (0L, 1), (1L, 1), (0L, 1), (0L, 1),
(1L, 1), (0L, 1), (1L, 1), (0L, 1), (1L, 1), (0L, 1), (1L, 1), (0L, 1),
(0L, 0), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1),
(0L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (0L, 0), (0L, 1),
(1L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (0L, 1), (1L, 1), (1L, 1),
(1L, 1), (0L, 0), (1L, 1), (0L, 0), (0L, 0), (1L, 1), (1L, 1), (1L, 1),
(0L, 1), (0L, 1), (1L, 1)]

[K-Nearest Neighbor]
Accuracy with K-NN: 82.3529411765%
[(0L, 0), (1L, 1), (1L, 1), (0L, 0), (1L, 0), (1L, 1), (1L, 1), (0L, 1),
(0L, 0), (0L, 0), (1L, 1), (1L, 1), (1L, 1), (0L, 1), (1L, 1), (0L, 0),
(1L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (0L, 0), (1L, 1), (1L, 1),
(0L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (1L, 1), (0L, 0), (0L, 1),
(1L, 1), (1L, 1), (0L, 0), (0L, 0), (0L, 1), (0L, 1), (1L, 1), (1L, 1),
(0L, 0), (0L, 0), (1L, 1), (1L, 1), (0L, 1), (1L, 1), (0L, 0), (0L, 1),
(1L, 1), (0L, 0), (0L, 0)]
```

Gambar 4.6 Implementasi Tampilan Sistem

BAB 5 PENGUJIAN DAN ANALISIS

Pada bab ini berisi tentang bagaimana proses pengujian berdasarkan hasil dari implementasi yang telah dijalankan pada program yang telah dibuat. Hasil dari implementasi pengujian antara lain, pengujian berdasarkan nilai atribut *K*, pengujian berdasarkan rasio perbandingan atau *percentage split* dan pengujian berdasarkan variasi data *training.* Bab ini juga berisikan analisis hasil pengujian berdasarkan nilai atribut *K,* analisis hasil pengujian berdasarkan rasio perbandingan data *training* dan data *testing* dan analisis hasil pengujian berdasarkan jumlah data *training.*

5.1 Hasil Pengujian

Pada sub bab ini berisi hasil pengujian yang dilakukan pada sistem berdasarkan pada pengujian yang telah dijelaskan pada bab sebelumnya yaitu pengujian berdasarkan nilai atribut *K* pada metode *K-Nearest Neighbor*, pengujian berdasarkan rasio perbandingan atau *percentage split* dan pengujian berdasarkan jumlah data training.

5.1.1 Hasil Pengujian Berdasarkan Nilai Atribut *K* pada Metode *K-Nearest Neighbor*

Pengujian ini dilakukan untuk mengetahui pengaruh nilai atribut *K* terhadap metode klasifikasi *K-Nearest Neighbor.* Pada pengujian ini dilakukan sebanyak 10 kali mengubah nilai *K* dalam setiap uji coba dimana setiap skenario pengujian *K* dilakukan sebanyak 5 kali. Hasil dari pengujian dengan nilai *K* terbaik akan digunakan dalam pengujian komparasi metode *K-Nearest Neighbor* dan *Naïve* Bayes. Nilai *K* yang digunakan pada pengujian komparasi nanti hanya menggunakan angka terkecil dengan akurasi terbaik dikarenakan untuk mempermudah penghitungan komaparasi selanjutnya. Data *training* yang yang digunakan pada pengujian sejumlah 127 data dan data *testing* sejumlah 40 data. Hasil pengujian berdasarkan nilai atribut *K* dapat dilihat pada Tabel 5.1.

Tabel 5.1 Tabel Hasil Pengujian Berdasarkan Nilai Atribut K pada Metode K-Nearest Neighbor

K	Uji Coba Ke-	Akurasi (%)
1	1	78.05%
	2	78.05%
	3	80.48%
	4	78.05%
	5	78.05%
Rata-Rata		**78.53%**
2	1	85.36%
	2	85.36%
	3	85.36%
	4	80.48%
	5	78.04%
Rata-Rata		**82.92%**

3	1	90.24%
	2	92.68%
	3	90.24%
	4	90.24%
	5	90.24%
Rata-Rata		**90.73%**
4	1	87.8%
	2	87.8%
	3	87.8%
	4	90.24%
	5	82.92%
Rata-Rata		**87.31%**
5	1	87.8%
	2	85.36%
	3	85.36%
	4	85.36%
	5	85.36%
Rata-Rata		**85.85%**
6	1	80.48%
	2	82.92%
	3	78.04%
	4	80.48%
	5	78.04%
Rata-Rata		**79.99%**
7	1	78.04%
	2	78.04%
	3	80.48%
	4	75.6%
	5	78.04%
Rata-Rata		**78.04%**
8	1	73.17%
	2	75.6%
	3	75.6%
	4	75.6%
	5	756%
Rata-Rata		**75.11%**
9	1	73.17%
	2	75.6%
	3	75.6%
	4	73.17%
	5	75.6%
Rata-Rata		**74.63%**

10	1	70.73%
	2	68.29%
	3	70.73%
	4	70.73%
	5	70.73%
Rata-Rata		**70.24%**

Berdasarkan hasil dari Tabel 5.1 dapat dilihat pada atribut *K* bernilai 10 menghasilkan nilai rata-rata akurasi yang paling rendah yaitu 70.24%. Nilai rata-rata akurasi terbaik terdapat pada pengujian dengan *K* bernilai 3 yaitu sebesar 90.73%. Dari hasil pengujian tersebut maka dapat ditarik kesimpulan akan menggunakan nilai atribut *K* = 3 dalam pengujian komparasi metode *K-Nearest Neighbor* dan *Nïave Bayes*.

5.1.2 Hasil Pengujian Berdasarkan Rasio Perbandingan atau *Percentage Split*

Hasil pengujian berdasarkan rasio perbandingan atau *percentage split* menggunakan data sejumlah 100% dari keseluruhan data atau sejumlah 167 data yang akan dibagi berdasarkan rasio perbandingan yang ditentukan yaitu menggunakan 90% data *training* dan 10% data *testing*, 80% data *training* dan 20% data *testing,* 70% data *training* dan 30% data *testing*, 60% data *training* dan 40% data. Dengan menggunakain nilai atribut *K* = 3. Pengujian ini dilakukan untuk mengetahui pengaruh rasio atau persentase tertentu dari jumlah data *training* dan data testing terhadap tingkat akurasi metode *K-Nearest Neighbor* dengan metode *Naïve Bayes.* Hasil pengujian dapat ditunjukan pada Tabel 5.2

Tabel 5.2 Tabel Hasil Pengujian Berdasarkan Rasio Perbandingan

Data *Training*	Data *Testing*	Uji Coba Ke-	Akurasi *k-nearest neighbor (%)*	Akurasi *naïve bayes (%)*
60% (99 data)	40% (68 data)	1	85.07%	70.15%
		2	79.1%	68.66%
		3	76.12%	73.13%
		4	76.12%	74.62%
		5	80.6%	68.66%
Rata-Rata			**79.40%**	**71.04%**
70% (116 data)	30% (51 data)	1	82.35%	76.47%
		2	80.4%	72.55%
		3	78.43%	68.63%
		4	80.4%	80.4%
		5	88.24%	64.71%
Rata-Rata			**81.96%**	**72.55%**
80% (133 data)	20% (34 data)	1	91.18%	79.41%
		2	82.35%	79.41%
		3	76.47%	64.71%
		4	80.4%	72.55%

		5	88.24%	68.63%
Rata-Rata			**83.73%**	**72.94%**
90% (150 data)	10% (17 data)	1	82.35%	76.47%
		2	88.24%	76.47%
		3	82.35%	64.71%
		4	82.35%	79.41%
		5	88.24%	70.59%
Rata-Rata			**84.71%**	**73,53%**
Rata-Rata Total			82.45%	72.52%

Berdasarkan hasil pengujian yang telah dilakukan, pada skenario pengujian dengan rasio perbandingan 90% data *training* dan 10% data testing metode *K-Nearest Neighbor* memiliki rata-rata akurasi paling tinggi yaitu hanya sebesar 84.71%. Sedangkan untuk metode *Naïve Bayes* memiliki nilai rata-rata akurasi paling tinggi juga pada skenario rasio perbandingan 90% data *training* dan 10% data *testing* yaitu sebesar 73.53%. Namun pada skenario rasio perbandingan 60% data *training* dan 40% data *testing*, metode *K-Nearest Neighbor* memiliki rata-rata nilai akurasi yang paling rendah yaitu 79.40%. Sama halnya dengan metode *Naïve Bayes* pada skenario rasio perbandingan 60% data *training* dan 40% data *testing* memiliki rata-rata nilai akurasi yang paling rendah yaitu sebesar 71.04% pada skenario tersebut.

5.1.3 Hasil Pengujian Berdasarkan Jumlah Data *Training*

Hasil pengujian berdasarkan variasi jumlah data *training* ini berbeda dengan pengujian sebelumnya yaitu *percentage split* dimana pada pengujian sebelumnya data yang digunakan 100% dari keseluruhan data. Pada pengujian ini tidak menggunakan data seluruhnya, hanya menggunakan beberapa data yang nantinya dibagi menjadi beberapa data *training* dengan jumlah data *training* yaitu 60 data, 80 data, 100 data, dan 120 data. Dan menggunakan jumlah data *testing* yang sama yaitu 40 data, hal ini disebabkan pengujian ini hanya difokuskan terhadap jumlah data *training*. Dengan menggunakan nilai atribut *K* yang sama dengan pengujian sebelumnya yakni 3. Pengujian ini bertujuan untuk mengetahui pengaruh jumlah data *training* terhadap tingkat akurasi yang dihasilkan tanpa melihat data *testing* pada metode *K-Nearest Neighbor* dan metode *Naïve Bayes*. Hasil pengujian berdasarkan variasi jumlah data *training* dapat dilihat seperti pada Tabel 5.3.

Tabel 5.3 Hasil Pengujian Berdasarkan Variasi Jumlah Data Training

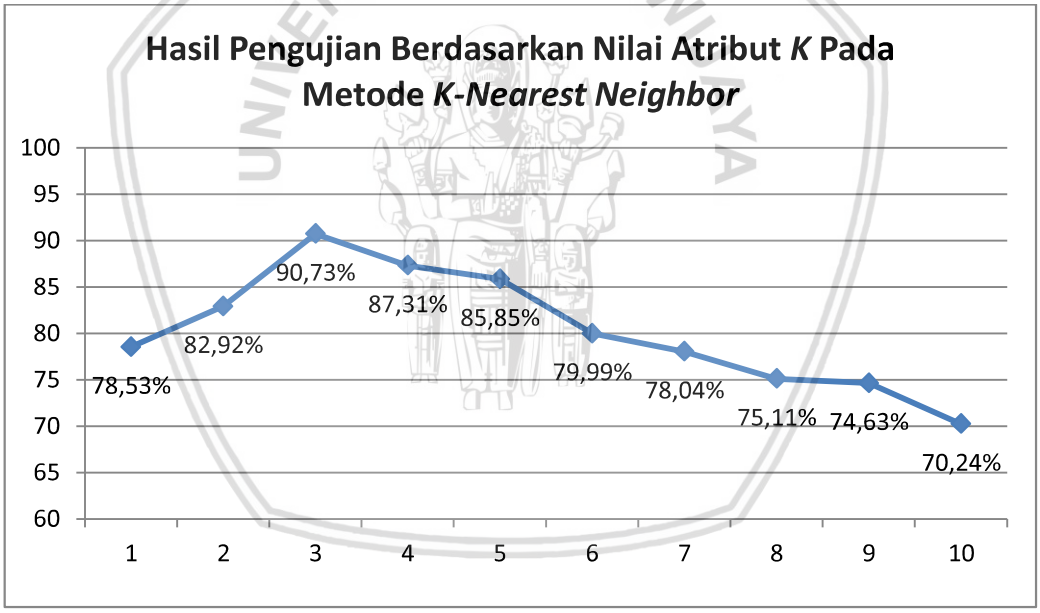
Data *Training*	Data *Testing*	Uji Coba Ke-	Akurasi *k-nearest neighbor (%)*	Akurasi *naïve bayes (%)*
60	40	1	85.37%	70,73%
		2	85.37%	70.73%
		3	75.6%	68.29%
		4	78.05%	63.41%
		5	85.37%	65.85%
Rata-Rata			**81,1%**	**67,07%**
80	40	1	80.49%	73.17%
		2	87.8%	78.05%
		3	80.49%	63.41%
		4	82.93%	80.49%
		5	85.37%	63.41%
Rata-Rata			**83,42%**	**71,72%**
100	40	1	87.8%	65.85%
		2	87.8%	75.6%
		3	82.93%	80.49%
		4	80.49%	73.17%
		5	80.49%	65.85%
Rata-Rata			**83,9%**	**72,19%**
120	40	1	90.24%	78.05%
		2	85.37%	70.73%
		3	82.92%	65.85%
		4	82.92%	70.73%
		5	82.92%	78.05%
Rata-Rata			**84,87%**	**72,67%**
Rata-Rata Total			**83,32%**	**70,91%**

Berdasarkan hasil pengujian yang telah dilakukan, metode *K-Nearest Neighbor* memiliki rata-rata nilai akurasi paling rendah pada pengujian dengan data *training* sejumlah 60 data yakni sebesar 81.1% dan rata-rata nilai akurasi paling tinggi dengan data *training* sejumlah 120 data dengan tingkat akurasi 84.87%. Pada metode *Naïve Bayes* rata-rata nilai akurasi paling rendah juga didapat pada pengujian dengan menggunakan data *training* sejumlah 60 dengan akurasi 67.07% dan rata-rata nilai akurasi paling tinggi sebesar 72.67% pada pengujian dengan data *training* sejumlah 120 data.

5.2 Analisis Hasil

5.2.1 Analisis Hasil Pengujian Berdasarkan Nilai Atribut *K* pada Metode *K-Nearest Neighbor*

Hasil pengujian berdasarkan nilai atribut *K* pada metode *K-Nearest Neighbor* akan ditampilkan dalam bentuk grafik pada Gambar 5.1. Pada pengujian berdasarkan nilai atribut *K* data yang digunakan sejumlah 167 data yang terbagi menjadi 127 data *training* dan 40 data *testing.* Hasil pengujian terendah terjadi ketika nilai atribut *K* bernilai 10 yaitu 70.24%. Hasil pengujian paling tinggi dan terbaik ketika nilai atribut *K* bernilai 3, dimana pada pengujian nilai tertinggi sebesar 90.73%. Pada Gambar 5.1 dapat ditarik kesimpulan bahwa akurasi *K-Nearest Neighbor* akan dipengaruhi terhadap jumlah nilai *K*. Semin banyak nilai *K,* maka semakin rendah tingkat akurasi, hal ini disebabkan oleh atribut yang digunakan memiliki kemiripan yang banyak sehingga semakin banyak tetangga atau nilai *K* yang diambil semakin banyak data dari kelas yang lain ikut dijadikan pertimbangan keputusan. Pada pengujian yang dilakukan peneliti akurasi yang terbaik didapatkan ketika *K* bernilai 3 yang nantinya akan digunakan dalam pengujian komparasi metode *K-Nearest Neighbor* dan *Naïve Bayes*



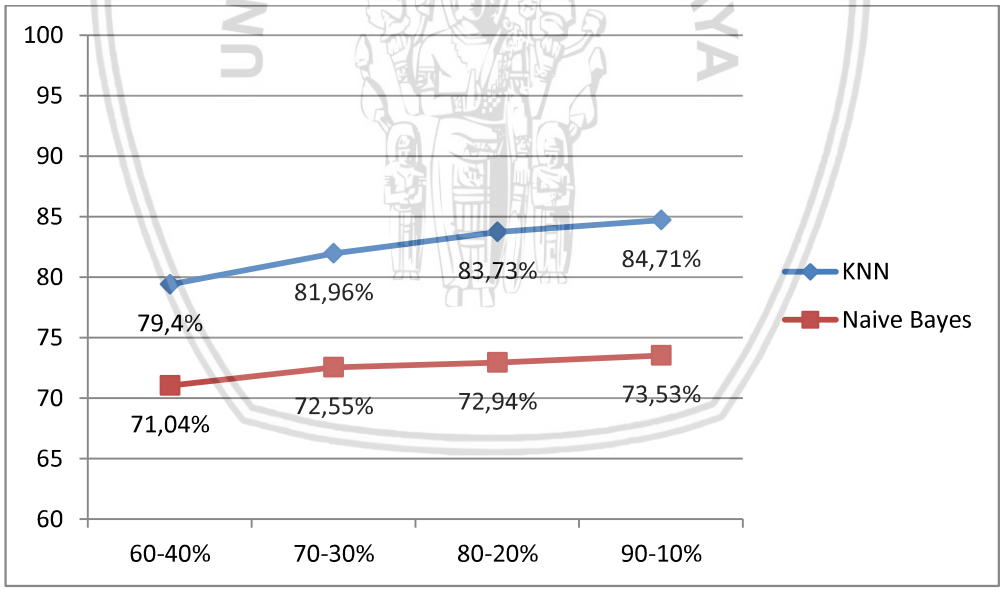
Gambar 5.1 Grafik Hasil Pengujian Berdasarkan Nilai Atribut K pada Metode K-Nearest Neighbor

5.2.2 Analisis Hasil Pengujian Berdasarkan Rasio Perbandingan atau *Percentage Split*

Pada pengujian ini data yang digunakan sejumlah 167 data yang terbagi berdasarkan rasio perbandingan data *training* dan data *testing* yang ditentukan yaitu menggunakan 90% data *training* dan 10% data *testing*, 80% data *training* dan 20% data *testing,* 70% data *training* dan 30% data *testing*, dan 60% data *training* dan 40% data *testing.* Nilai atribut *K* yang digunakan bernilai 3. Hasil pengujian metode *K-Nearest Neighbor* menghasilkan nilai rata-rata akurasi paling

rendah pada rasio perbandingan 60% data *training* dan 40% data *testing* yaitu dengan akurasi 79.4%. Sedangkan metode *Naïve Bayes* juga memiliki rata-rata nilai akurasi paling rendah pada rasio perbandingan 60% data *training* dan 40% data *testing* yaitu dengan akurasi 71.04%. Pada metode *Naïve Bayes* memiliki rata-rata nilai akurasi paling tinggi yaitu 73.53% pada rasio perbandingan 90% data *training* dan 10% data *testing*. Sedangkan rata-rata nilai akurasi paling tinggi dari metode *K-Nearest Neighbor* juga terdapat pada rasio perbandingan 90% data *training* dan 10% data *testing* yaitu sebesar 84.71%. Nilai rata-rata keseluruhan akurasi metode *K-Nearest Neighbor* yakni sebesar 82.45% dan nilai rata-rata keseluruhan akurasi metode *Naïve Bayes* sebesar 72.52%. Hasil pengujian berdasarkan rasio perbandingan data *training* dan data *testing* akan ditampilkan pada grafik pada Gambar 5.2.

Pada grafik pada Gambar 5.2 diperlihatkan bahwa semakin besar selisih persentase atau rasio antara data *training* dan data *testing* maka semakin tinggi pula akurasi yang didapatkan. Hal tersebut dikarenakan jumlah data *training* yang semakin lebih banyak daripada jumlah data *testing* yang semakin lebih sedikit maka model *classifier* yang dibangun berdasarkan fakta dari data *training* akan lebih baik dan lebih lengkap untuk melakukan prediksi terhadap klasifikasi data baru atau data *testing*. Sehingga akurasi yang didapatkan akan jauh lebih baik juga.

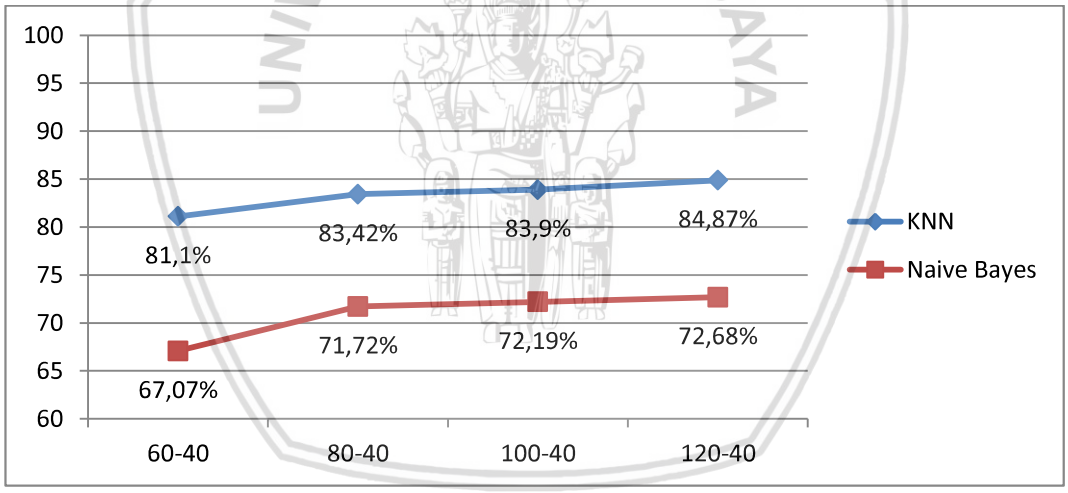


Gambar 5.2 Grafik Hasil Pengujian Berdasarkan Rasio Perbandingan

5.2.3 Analisis Hasil Pengujian Berdasarkan Jumlah Data *Training*

Hasil pengujian berdasarkan variasi jumlah data *training* akan ditampilkan pada grafik seperti pada Gambar 5.3. Hasil pengujian metode *K-Nearest Neighbor* memperoleh rata-rata nilai akurasi tertinggi sebesar 84.87% dengan data *training* sebanyak 120 data dan rata-rata nilai akurasi terendah sebesar 81.1% dengan jumlah data 60 data *training*. Rata-rata akurasi dari metode *K-Nearest Neighbor* adalah 83.32%. Sedangkan hasil pengujian metode *Naïve Bayes* diperoleh rata-rata nilai akurasi tertinggi sebesar 72.68% dengan data *training* sebanyak 120 data dan rata-rata nilai akurasi terendah sebesar 67.07% dengan jumlah data 60 data *training*. Rata-rata akurasi dari metode *Naïve bayes* adalah 70.91%. Pada grafik Gambar 5.3 diketahui bahwa semakin banyak jumah data *training* semakin tinggi pula tingkat akurasi metode.

Sama halnya dengan pengujian berdasarkan rasio perbandngan atau *percentage split*, semakin banyak jumlah data *training* maka semakin tinggi pula hasil akurasi yang didapatkan. Hal tersebut dikarenakan jika semakin banyak data *training* yang digunakan maka semakin baik dan semakin lengkap juga model *classifier* yang dibentuk berdasarkan fakta dari data *training* tersebut. Sehingga pada saat melakukan prediksi pada klasifikasi data *testing* atau data baru, akurasi yang didapatkan akan semakin baik atau tinggi.



Gambar 5.3 Grafik Hasil Pengujian Berdasarkan Jumlah Data Training

BAB 6 PENUTUP

6.1 Kesimpulan

Berdasarkan perancangan, implementasi dan pengujian dari sistem yang telah dilakukan oleh peneliti, maka dapat ditarik kesimpulan bahwa :

1. Implementasi algoritme *K-Nearest Neighbor* dan *Naive Bayes* pada klasifikasi kualitas air bersih, atribu-atribut yang digunakan dalam mebangun sistem berupa atribut komposisi pada air bersih meliputi *Coliform, Escherichia Coli,* Mangan, TDS dan Khlorida. Data-data yang diperoleh peneliti berasal dari PDAM Tirta Kencana Kabupaten Jombang sejumlah 167 data yang mana data tersebut dibagi menjadi 127 data *training* dan 40 data *testing*.
   Langkah selanjutnya untuk masing-masing metode sebagai berikut :
   - Pada metode *K-Nearest Neighbor,* langkah pertama yakni menghitung jarak *Euclidean*, setelah perhitungan jarak *Euclidean* data diurutkan berdasarkan nilai jarak terkecil, kemudian diambil data dengan jarak terdekat sebanyak nilai *K.* Setelah itu dapat dilihat kelas mana yang kemunculannya paling banyak , maka kelas tersebut merupakan hasil dari keputusan sistem
   - Pada metode *Naïve Bayes,* langkah pertama yakni melakukan perhitungan *prior* untuk masing-masing atribut pada air bersih. Setelah itu juga dilakukan perhitungan probabilitas *likelihood* juga terhadap masing-masing nilai atribut pada air bersih. Kemudian sistem akan menghitung nilai *posterior* untuk mengambil hasil keputusan sistem berdasarkan nilai probabilitas *posterior* yang paling besar

2. Pada penelitian ini, baik metode *K-Nearest Neighbor* dan metode *Naïve Bayes* memiliki akurasi sebagai berikut :
   - Pada pengujian berdasarkan nilai atribut *K* pada metode *K-Nearest Neighbor* didapatkan rata-rata nilai *K* yang terbaik yaitu *K* = 3 yang mana nilai tersebut merupakan nilai *K* yang memiliki rata-rata tingkat akurasi paling baik yaitu 90.73%. Nilai *K* tersebut digunakan dalam pengujian komparasi metode berdasarkan rasio perbandingan atau *percentage split* dan berdasarkan jumlah data *training*
   - Rata-rata nilai akurasi tertinggi metode *K-Nearest Neighbor* pada pengujian berdasarkan rasio perbandingan atau *percentage split* yaitu sebesar 84.71% pada skenario pengujian dengan rasio 90% data *training* dan 10% data *testing*. Rata-rata keseluruhan nilai akurasi metode *K-Nearest Neighbor* adalah 82.45%. Berdasarkan jumlah data *training* yaitu sebesar 84.87% pada skenario pengujian dengan data *training* sejumlah 120 data. Rata-rata keseluruhan nilai akurasi metode *K-Nearest Neighbor* adalah 83.32%
   - Rata-rata nilai akurasi tertinggi metode *Naïve Bayes* pada pengujian berdasarkan rasio perbandingan atau *percentage split* yaitu sebesar 73.53% pada skenario pengujian dengan rasio 90% data *training* dan 10% data *testing*. Rata-rata keseluruhan nilai akurasi metode *Naïve Bayes* adalah 72.52%. Berdasarkan jumlah data *training* yaitu sebesar 72.68%

pada skenario pengujian dengan data *training* sejumlah 120 data. Rata-rata nilai akurasi metode *Naïve Bayes* adalah 70.91%

- Berdasarkan hasil akurasi yang dihasilkan dari seluruh pengujian dapat disimpulkan metode terbaik yang sesuai dengan klasifikasi kualitas air bersih adalah metode *K-Nearest Neighbor* dengan rata-rata total akurasi sebesar 82.89%

3. Pengaruh nilai *K*, rasio perbandingan data *training* dan data *testing* dan variasi jumlah data *training* terhadap akurasi metode *K-Nearest Neighbor* dan *Naïve Bayes*:
   - Pengaruh nilai *K* terhadap akurasi pada algoritme *K-Nearest Neighbor* yaitu apabila nilai *K* semakin banyak maka akurasi yang didapat semakin kecil, hal tersebut dikarenakan oleh semakin banyak tetangga atau nilai *K* yang diambil semakin banyak data ikut dijadikan pertimbangan keputusan kelas.
   - Rasio perbandingan jumlah data *training* dan data *testing* juga memiliki pengaruh terhadap akurasi algortime *K-Nearest Neighbor* dan *Naïve Bayes,* semakin besar selisih persentase atau rasio antara data *training* dan data *testing* maka semakin tinggi pula akurasi yang didapatkan. Hal tersebut dikarenakan jumlah data *training* yang semakin lebih banyak daripada jumlah data *testing* yang semakin lebih sedikit maka model *classifier* yang dibangun berdasarkan fakta dari data *training* akan lebih baik dan lebih lengkap untuk melakukan prediksi terhadap klasifikasi data baru atau data *testing*. Sehingga akurasi yang didapatkan akan jauh lebih baik juga.
   - Perbedaan jumlah data *training* juga memiliki pengaruh terhadap akurasi algoritme *K-Nearest Neighbor* dan *Naive Bayes.* Sama halnya dengan pengujian berdasarkan rasio perbandngan atau *percentage split*, semakin banyak jumlah data *training* maka semakin tinggi pula hasil akurasi yang didapatkan. Hal tersebut dikarenakan jika semakin banyak data *training* yang digunakan maka semakin baik dan semakin lengkap juga model *classifier* yang dibentuk berdasarkan fakta dari data *training* tersebut. Sehingga pada saat melakukan prediksi pada klasifikasi data *testing* atau data baru, akurasi yang didapatkan akan semakin baik atau tinggi.

6.2 Saran

Klasifikasi kualitas air bersih masih memilki banyak kekurangan dan jauh dari kata sempurna. Berdasarkan hasil penelitian yang telah dilakukan penulis, terdapat beberapa saran yang diberikan untuk pengembangan penelitian dan sistem agar menjadi lebih baik antara lain :

1. Menggunakan metode yang lain selain dengan *K-nearest Neighbor* dan *Naïve Bayes* dalam perbandingan metode klasifikasi.
2. Menambahkan parameter dan metode dalam pengklasifikasian kualitas air bersih agar hasil yang didapatkan lebih bervariasi.

DAFTAR PUSTAKA

Bustami. 2013. "*Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*". TECHSI : Jurnal Penelitian Teknik Informatika, Vol. 3, No.2, Hal. 127-146

Donna Prayoga, Novianto. 2018. "*Sistem Diagnosis Penyakit Hati Menggunakan Metode Naive Bayes".* Universitas Brawijaya. Malang.

Hamidi, Rifwan,dkk. 2017. "*Implementasi Learning Vector Quantization (LVQ) untuk Klasifikasi Kualitas Air Sungai*". Universitas Brawijaya, Malang.

Hastuti, Khafiizh. 2012. "*Analisis Komparasi Algoritme Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif*". Universitas Dian Nuswantoro, Semarang.

Iskandar, Derick dan Suprapto, Yoyon K. 2013. "*Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan antara Algoritme C4.5 dan Naive Bayes Classifier*". Institut Teknologi Sepuluh November, Surabaya.

J. Kodoatie, Robert & Roestam Sjarief. 2010. "*Tata Ruang Air*". Yogyakarta.

Kusnawi. 2007. "*Pengantar Solusi Data Mining*". STMIK AMIKOM, Yogyakarta.

Kustiyahningsih Yeni, dkk. 2013. "*Sistem Pendukung Keputusan untuk Menentukan Jurusan pada Siswa SMA Menggunakan Metode KNN dan SMART*". Universitas Trunojoyo, Madura.

Kusumadewi, Sri. 2009. "*Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification*". Universitas Islam Indonesia, Yogyakarta.

Lestiana, Mila. 2015. "*Perbandingan Algoritma Decision Tree (C4.5) dan Naive Bayes pada Data Mining untuk Identifikasi Tumbuh Kembang Anak Balita*". Universitas Muhammadiyah, Surakarta.

Peraturan Menteri Republik Indonesia nomor 492 tahun 2010 tentang Persyaratan Kualitas Air minum. Jakarta : Kementrian Kesehatan Republik Indonesia.

Peraturan Pemerintah Republik Indonesia nomor 82 tahun 2001 tentang Pengelolaan Kualitas Air dan Pengendalian Pencemaran Air. Jakarta.

Santoso. 2016. "*Perbandingan Metode K-Nearest Neighbor(K-NN) dan Learning Vector Quantization (LVQ) untuk Permasalahan Klasifikasi Tingkat Kemiskinan".* Institut Teknologi Sepuluh November, Surabaya.

Situmorang, Manihar. 2017. "*Kimia Lingkungan*". Depok: PT RajaGrafindo

Tri Vulandari, Retno. 2017. "*Data Mining Teori dan Aplikasi Rapidminer*". Surakarta.